INCERTAE SEDIS NO MORE: THE PHYLOGENETIC AFFINITY OF HELICOSPORIDIA

By

AURELIEN TARTAR

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2004

To my wife, Jaime

ACKNOWLEDGMENTS

During my doctoral studies at the University of Florida, I have met diverse and numerous people that contributed in refining my scientific work and judgment, and I am thankful to all of them. I would like to express my deepest appreciation to my graduate committee chair, Dr. Drion Boucias, for welcoming me in his home and his laboratory, guiding and supporting me while allowing me to mature as an independent scientist and human being. I have no doubt that Drion is a unique mentor and a gifted scientist, and he will remain both my model and my friend. I would like to extend thanks to his wife and his family.

I am similarly grateful to the remaining members of the graduate committee, Drs. James Maruniak, Byron Adams, William Farmerie and Dave Clark, for the time, help, support, guidance, critical reviews and additional expertise they provided. They all have contributed in broadening my knowledge and interests and in increasing my conviction that remarkable mentors have surrounded me throughout my doctoral studies.

I would also like to thank Dr. James Becnel, Dr. Sasha Shapiro, and Susan White for providing me with the opportunity to work on the *Helicosporidium* isolates that they collected and expressing their support and encouragement in each of our regular meetings.

I thank Dr. Patrick Keeling at the University of British Columbia for initiating our collaborative EST project. Patrick, and his student Audrey, allowed my work to be more

complete and demonstrated an interest in my research that provided me with great support and confidence.

I would like to acknowledge the financial support provided by the National Science Foundation, as well as the various organizations and professional societies that, through grant support, allowed me to present my work around the world.

Finally, I will be forever grateful for the molecular techniques class offered by the Interdisciplinary Center for Biotechnology Research in July/August 2001. My lab mate for this class, Jaime, has become the most important person in my life, my wife.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure page

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

INCERTAE SEDIS NO MORE:
THE PHYLOGENETIC AFFINITY OF HELICOSPORIDIA

By

Aurélien Tartar

May 2004

Chair: Drion G. Boucias
Major Department: Entomology and Nematology

The Helicosporidia are a unique group of pathogens found in diverse invertebrate hosts. They have been considered to be either protozoa or fungi but have remained *incertae sedis* since 1931. Following the isolation of a new *Helicosporidium* sp. in Florida, the Helicosporidia were characterized as non-photosynthetic green algae (Chlorophyta). Phylogeny reconstructions inferred on several housekeeping genes (including actin and β-tubulin) consistently and stably grouped *Helicosporidium* sp. among members of Chlorophyta. Additionally, nuclear SSU rDNA phylogenies identified *Helicosporidium* as a sister taxon to another parasitic, non-photosynthetic algal genus: *Prototheca* (Chlorophyta, Trebouxiophyceae). Comparison of mitochondrial (*cox3*) and chloroplast (*rrn16*) genes confirmed that *Helicosporidium* and *Prototheca* have arisen from a common photosynthetic ancestor and suggested that Helicosporidia contain *Prototheca*-like organelles, including a vestigial chloroplast (plastid). A fragment of the *Helicosporidium* sp. plastid DNA (ptDNA) has been amplified and sequenced.

Comparative genomic analyses, coupled with RT-PCR amplifications performed on the ptDNA fragment, demonstrated that *Helicosporidium* sp. has retained a modified but functional plastid genome. In addition, the Helicosporidia were shown to possess a reduced nuclear genome. Lastly, in an effort to better characterize the biology of *Helicosporidium* sp., a cDNA library has been constructed and expressed sequences tags (ESTs) have been generated. Most of these ESTs exhibited similarity to algal and plant genes, and additional phylogenetic analyses inferred from selected ESTs confirmed the green algal nature of *Helicosporidium* sp. The EST database provides insights into the biology and the evolution of the Helicosporidia. Notably, the sequencing of a bacterial protease from the *Helicosporidium* sp. genome suggests that the Helicosporidia may have acquired virulence factors via lateral gene transfer from an unrelated organism. Overall, the data accumulated throughout this study are all concordant with the conclusion that the Helicosporidia are highly adapted, non-photosynthetic, parasitic green algae.

CHAPTER 1
INTRODUCTION AND RESEARCH OBJECTIVES

The Helicosporidia are a unique group of pathogens that have been detected in a variety of invertebrate hosts. Like other insect pathogens, the Helicosporidia have been studied because of their potential as biocontrol agents. However, they remain little-known organisms, and, to date, their importance and occurrence as invertebrate pathogens are unclear. Notably, their taxonomic status has remained *incertae sedis*, meaning that it has not been finalized. Because of its uncertain evolutionary affinity, most recent reviews of insect pathogens hardly mention the group's existence (Tanada and Kaya, 1993; Undeen and Vavra, 1997), or ignore it (Boucias and Pendland, 1998), and only a handful of scientific reports have been published on these organisms.

Literature Review of *Helicosporidium* spp.

To date, there is only one named species of Helicosporidia: *Helicosporidium parasiticum*. It was initially described and named by Keilin (1921), who detected this protist in larvae of *Dasyhelea obscura* Winnertz (Diptera: Ceratopogonidae) collected in England. He examined the new parasite thoroughly and attempted to infer its life history from his observations. He characterized a vegetative growth by very active multiplications of helicosporidial cells within the host hemocoel and noticed that these "schizogonic multiplications" were followed by the formation of what he called spores. Keilin noted that the spores were very easily recognized: they consisted of three ovoid cells (named by Keilin "sporozoites") and one peripheral, spiral, filamentous cell, assembled inside an external membrane. These features, especially the highly

characteristic filamentous cell, have since remained the principal diagnostic for identification of a *Helicosporidium* sp. Keilin was able to describe and characterize structurally the new genus *Helicosporidium* and the new species *H. parasiticum*. He was also able to present a hypothetical life cycle of this protist based on microscopic observations. He suggested that the spores (or cysts) break open in the host hemocoel, releasing the filamentous cell and the three "sporozoites," which he proposed are the infective forms of *H. parasiticum*. He also provided information on frequency of infection and on potential new host species for this pathogen, including the dipteran *Mycetobia pallipes* Meig. and the mite *Hericia hericia* Kramer (Keilin, 1921).

Despite all the data gathered on this organism, Keilin was not able to answer the question of the systematic position of *Helicosporidium parasiticum*. He believed that *H. parasiticum* belonged to the Protozoa, and he compared his isolate with members of various clades: Cnidiosporidia (which, at that time, included Microsporidia such as *Nosema bombicis*), Haplosporidia, Serumsporidia, and Mycetozoa. He concluded that the genus *Helicosporidium* differed markedly not only from all these groups, but also from all the protists known at that time. He finally proposed that Helicosporidium "forms a new group, which may be temporarily included in the group of the Sporozoa" (Keilin, 1921, p. 110).

Kudo (1931) was the first one to associate the genus *Helicosporidium* with other known organisms. He considered that *Helicosporidium parasiticum* was a protozoan, and, based on Keilin's description, placed it within the Cnidosporidia in a separate order that he created and named Helicosporidia. In his classification, the closest group to Helicosporidia was the order Microsporidia.

Following the discovery of another isolate of *Helicosporidium parasiticum* in a larva of *Hepialis pallens* (Hepialidae, Lepidoptera), another taxonomic position was proposed for the group Helicosporidia (Weiser, 1970). Based on observation of this new isolate as well as the original specimen described by Keilin, Weiser claimed that the Helicosporidia were best placed among the lower Fungi. He argued that the spore characteristics were much too different from what was found in Protozoa, but they were similar in some aspects to primitive Fungi, such as insect pathogens of the genus *Monosporella*, classified as Nematosporoideae inside the Saccharomycetaceae (primitive Ascomycetes).

Kellen and Lindegren (1973) reported the third isolation of *Helicosporidium parasiticum*, this time from larvae and adults of the beetle *Carpophilus mutilatus* (Nitidulidae, Coleoptera). With this isolate, they successfully infected *per os* 18 species of arthropods belonging to three orders of insects (Lepidoptera, Coleoptera, Diptera) and one family of mites. They also were able to note that some species of Orthoptera, Hymenoptera, and Diptera are not susceptible to their isolates. Their report is the first host range study for an isolate of *Helicosporidium parasiticum*. Importantly, they used their isolate to infect larvae of the navel orangeworm *Paramyelois transitella* (Phyralidae, Lepidoptera), which were easily manipulated in the laboratory, and used this host/pathogen model to study the life cycle of *H. parasiticum* (Kellen and Lindegren, 1974). This led them to detail a *Helicosporidium* life cycle that differed from the one proposed by Keilin. They observed that *H. parasiticum* is infectious *per os*. The spores, present in the host artificial diet, were ingested and released the three round cells and the filamentous cells in the host midgut. After 24h, helicosporidial cells appeared in the host

hemolymph and grew vegetatively. The vegetative growth was characterized by cell division that occured within a pellicle. After division, the pellicle ruptured and released the daughter cells (4 or 8). Empty pellicles and daughter cells eventually filled the entire host hemocoel. Daughter cells then developed into spores in which the filamentous cell differentiated and encircled the three round cells. These observations allowed Kellen and Lindegren to better characterize the infectious process of *Helicosporidium parasiticum* in a lepidopteran host. Their knowledge led them to express doubt about the validity of Weiser's taxonomic classification. They proposed that the group Helicosporidia should be removed from the Protozoa, as Weiser (1970) proposed, but they also argued that this group was not closer to the Fungi than it was to the Protozoa. However, they were unable to suggest a better classification.

Later work by Lindegren and Hoffman (1976) and Fukuda et al. (1976) added yet more confusion about the Helicosporidia as a group. First, ultrastructure studies, based on transmission electron microscopy (TEM) pictures of various developmental stages of the *Helicosporidium parasiticum* isolated from the beetle, led Lindegren and Hoffman (1976) to conclude that the Helicosporidia are related to the Protozoa. Their conclusion was based on the presence of well-defined Golgi bodies and observations of mitotic division of the nucleus. Additionally, Lindegren and Hoffman (1976) compared their *Helicosporidium* isolate to another one isolated from a mosquito larva of *Culex territans*. They noted that these two isolates resembled one another more than any resembled the original isolate described by Keilin. Thus, they introduced the hypothesis that there may be more than one species of *Helicosporidium*. Consequently, when they reported the

isolation of their novel *Helicosporidium* sp. isolate, Fukuda et al. (1976) referred to both isolates as the "beetle *Helicosporidium*" and the "mosquito *Helicosporidium*."

After Lindegren and Hoffman (1976) had proposed that the Helicosporidia have affinities to the Protozoa, the debate about the taxonomic position of Helicosporidia terminated. However, Lindegren and Hoffman (1976) failed to associate the Helicosporidia with any known protozoan group, and they proposed additional taxonomic studies. These have never happened. The subsequent studies on various *Helicosporidium* isolates consist, for the most part, of reports of the presence of *Helicosporidium* sp. in new host species, such as crustaceans, mites and collembola, trematodes, or even free-living forms of *Helicosporidium* sp. (Sayre and Clarke, 1978; Hembree, 1979, 1981; Purrini, 1984; Kim and Avery, 1986; Avery and Undeen, 1987a, b; Pekkarinen, 1993; Seif and Rifaat, 2001). Most of these studies refer to the Helicosporidia as a subphylum of Protozoa, and have little mention of their potential phylogenetic affinities. The spelling of the original order created by Kudo (1931) even suffered and became "Helicosporida," with no apparent reasons or explanations (see Sayre and Clarke, 1978; Hembree, 1979, 1981; Pekkarinen, 1993; Seif and Rifaat, 2001).

Therefore, the only attempted classification for the Helicosporidia is the one proposed in 1931 by Kudo, who placed this group as a close relative of Microsporidia in a subphylum (Cnidiospora) of Protozoa. Aside from this classification, the Helicosporidia have remained *incertae sedis*, or, at best, Protozoa *incertae sedis*. The group has never appeared in other taxonomic classifications, and it is absent from the most recent classification systems of either the Protozoa or the Fungi.

The Helicosporidia: More Than Ever *incertae sedis*

The classification of Helicosporidia as Protozoa *incertae sedis* reflects the fact that these organisms have never been related to any other known protist. As noted by Undeen and Vavra (1997), "the (helicosporidial) spores are characteristic and not easily mistaken for any other protozoan, particularly after they have been germinated or crushed under a coverslip, revealing the coiled filamentous cells." Nevertheless this taxonomy, or lack thereof, also reflects a poor knowledge of this group. It is all the more unsatisfactory that contemporary methods, such as molecular sequence comparative analyses, have contributed to improve the knowledge on eukaryote evolution, and have led to the identification of major eukaryotic groups. Being absent from most taxonomic classifications, the Helicosporidia have been ignored from the dramatic changes in understanding of eukaryotic phylogenies.

"Protozoa" Is an Obsolete Phylum

The tremendous progress in resolving deep eukaryotic taxonomy has been reviewed by several authors (Simpson and Roger, 2002; Baldauf, 2003; see also Cavalier-Smith and Chao, 2003). They present a relatively similar consensus phylogeny of eukaryotes obtained by the combination of evidence from molecular sequence trees, morphology, biochemistry, and discrete genetic characters such as indels and gene fusions that can be treated cladistically. The authors agree that, despite being clearer than ever, the general understanding of eukaryotic phylogeny is still improving, and there remain a number of major gaps, especially in regard to the relationships among eukaryote supergroups and the position of the root that would link eukaryotes and prokaryotes. These gaps explain the difference in numbers of supergroups reported by the different

reviews: Baldauf (2003) lists eight major groups, while Simpson and Roger (2002) sort eukaryotes into six groups.

In the most recent and conservative analysis (Bauldauf, 2003), eight supergroups are recognized: Opisthokonts (animals, fungi, choanoflagellates), Plants, Amoebozoa, Cercozoa (cercomonads, foraminiferans), Alveolates (dinoflagellates, ciliates, Apicomplexa), Heterokonts (a.k.a. Stramenopiles: brown algae, diatoms, oomycetes), Discicristates (kinetoplasts) and Excavates (diplomonads, parabaselids). Other analyses (i.e. Simpson and Roger, 2002) include the Discicristates in the Excavates and group the Alveolates and Heterokonts in one supergroup named Chromalveolates, leading to a six-group-based classification of eukaryotes which includes Opisthokonts, Plants, Amoebozoa, Cercozoa, Chromalveolates and Excavates. Most significantly, these two classifications are remarkably similar in that they fail to mention the phylum "Protozoa." Although the term "protozoa" is still used in some contemporary reviews, such as one by Cavalier-Smith and Chao (2003), it has become clear that this grouping of eukaryotes is not supported by recent molecular sequence-based phylogenies. Cavalier-Smith and Chao (2003) identify the "kingdom Protozoa" as a polyphyletic group divided into two infrakingdoms: the Alveolates (that are nonetheless classified within the supergroup Chromalveolates in the same study) and the Excavates. More data and improved methods are constantly accumulating and improving the resolution of these deep-branching supergroups and their relationships to each other, likely leading to the complete collapse of the "Protozoa" notion. This collapse is exemplified by the recent publication of *The Illustrated Guide to the Protozoa* 2nd Edition (Lee et al., 2002) which has been subtitled *Groups Classically Considered Protozoa and Newly Discovered Ones.*

Because they never have been related to any other known unicellular organisms, the Helicosporidia cannot be classified within any of the newly identified eukaryotic supergroups. Significantly, the group has never been subjected to contemporary molecular-sequence-based phylogenetic analyses that have accounted for much of this fundamental rethinking of eukaryotic evolution. In contrast, other (ex-)protozoan groups, such as the Microsporidia, which were proposed by Kudo (1931) to be the closest relatives to Helicosporidia, have been the subject of a complete taxonomic re-assignment.

Microsporidia Are Fungi

Microsporidia are obligate intracellular parasites of eukaryotes. The majority of the more than 1000 described species have been detected in insect hosts. Significantly, the first known microsporidium, *Nosema bombycis*, was identified by Louis Pasteur as the causal agent of the pebrine disease in the silkworm *Bombyx mori*. Microsporidia are identified by the production of small spores containing a polar filament that is involved in a highly specialized mode of infection. They are also characterized by the presence of a prokaryotic 70S ribosomal DNA and the lack of mitochondria. In addition, rDNA small subunit phylogenies placed the Microsporidia at a very basal position in the eukaryotic tree. As a result, these organisms were believed to be very primitive eukaryotes that may have diverged very early, possibly before the acquisition of mitochondria by other eukaryotes. However, molecular data, especially from protein-coding genes, have accumulated and, although some analyses remain contradictory (reviewed by Keeling and Fast, 2002), there are now a number of gene phylogenies that provide strong support for a Microsporidia-Fungi relationship. A recent analysis even suggested that Microsporidia are related to zygomycetes (Keeling, 2003). Furthermore, other types of evidence, such as

the discovery of relic mitochondrial genes in microsporidian genomes, have supported the hypothesis that Microsporidia are extremely modified and reduced fungi that have secondarily lost organelles such as mitochondria.

At different points in time, the Helicosporidia were proposed to be either close relatives to Microsporidia (Kudo, 1931) or to Fungi (Weiser, 1970). Interestingly, that ambiguity is somewhat concordant with the reclassification of Microsporidia as Fungi. However, as stated before, the Helicosporidia have never been included in any recent taxonomic revisions, including those involving the Microsporidia. Today, it is unclear whether this group should be re-associated with the Microsporidia, within the Fungi, or if it belongs to one of the newly identified eukaryotic supergroups or even forms a completely unique eukaryote taxon. The group remains, more than ever, *incertae sedis*.

New Findings on Helicosporidia

In 1999, a *Helicosporidium* sp. was discovered in larvae of the black fly *Simulium jonesi* Stone & Snoddy (Simuliidae, Diptera) collected in Gainesville, Florida (Boucias et al., 2001). The detection of this isolate and the ability to produce quantities of this pathogen in a laboratory insect such as *Helicoverpa zea* stimulated additional studies on Helicosporidia. The authors identified *Helicosporidium* sp. based on the highly characteristic cyst that encloses three ovoid cells and a spiral filamentous cell. They described this isolate using both light and electron microscopy, and they examined its life cycle and its infectious process in the laboratory insects *Helicoverpa zea*, *Manduca sexta*, and *Galleria mellonella*. They observed a very similar infectious pattern as previously reported. They showed that helicosporidial cysts are ingested by suitable hosts and that physicochemical conditions within the midgut stimulate cyst dehiscence. The ovoid cells

and the filamentous cells are then released, and the filamentous cells attach to the peritrophic membrane. According to Boucias et al. (2001), the three ovoid cells are short-lived in the insect gut, and infection is mediated by filamentous cells. The authors also performed some host range studies as well as some *in vitro* propagation experiments. Interestingly, they suggested that the vegetative growth of *Helicosporidium* sp. observed in artificial media was reminiscent of what has been reported for unicellular, achlorophytic algae belonging to the genus *Prototheca*. Both the genera *Helicosporidium* and *Prototheca* are characterized by a vegetative growth that consists of cell divisions inside a membrane. Four, eight, or sixteen daughter cells are produced inside this pellicle and are eventually released. Such cell divisions result in the accumulation of both round daughter cells and empty pellicles. Boucias et al. (2001) also noted that, like *Helicosporidium* spp., *Prototheca* spp. are pathogenic but have been associated solely with vertebrates. Furthermore, *Prototheca* spp. are not known to produce the filamentous cell-containing cyst, which is characteristic of the genus *Helicosporidium*. Finally, the authors expressed some doubt about the possible protozoan nature of Helicosporidia: they argued that *Helicosporidium* sp. has very simple growth requirements and can be cultivated in various artificial media. This characteristic made it very different from other known entomopathogenic organisms traditionally classified as Protozoa.

Research Objectives

The Helicosporidia is an enigmatic group that has been poorly studied. Although there are more and more data describing its potential hosts, general life cycle, and pathogenicity process, the general understanding of this unique genus is scant when compared to other entomopathogenic genera. In particular, its taxonomic status has

remained a mystery since its first discovery. The Helicosporidia have successively been associated with Protozoa, Fungi, or Algae, but they remain, despite these attempts, *incertae sedis*. Developing fundamental knowledge on the genus *Helicosporidium* may become more and more crucial, since these organisms recently have been examined as potential biocontrol agents against mosquitoes (Hembree, 1981; Kim and Avery, 1986; Avery and Undeen, 1987; Seif and Rifaat, 2001). Precisely determining the taxonomic position of *Helicosporidium* spp. within the eukaryotic tree will be an important step toward increasing knowledge of these organisms.

The overall objective of this project is to determine the position of the genus *Helicosporidium* within the eukaryotic tree of life and to associate these organisms with other known protists. Modern methods, such as comparative sequence analyses, will be used. Such methods have been shown to provide resolving power for clade identification. The study will focus on producing DNA sequence information from *Helicosporidium* sp. that can be used to inform taxonomic statements. One priority is to compare the Helicosporidia with the genus *Prototheca*, which has been identified as a potential close relative of *Helicosporidium* sp. by Boucias et al. (2001). I will use the *Helicosporidium* sp. isolate detected by these authors in a black fly larva collected in Florida, as it is now fully established in *in vitro* cultures, on artificial media, and has been shown to be suitable for DNA extraction and amplification (Boucias et al., 2001).

CHAPTER 2
NUCLEAR GENE PHYLOGENIES

Introduction

The Helicosporidia are a unique group of pathogens found in diverse invertebrate hosts. Members of this group are characterized by the formation of a cyst stage that contains a core of three ovoid cells and a single filamentous cell (Kellen and Lindegren, 1974; Lindegren and Hoffman, 1976). The group is very poorly known and its taxonomic position has remained *incertae sedis*. This pathogen, initially detected in a ceratopogonid (Diptera), was described and named *Helicosporidium parasiticum* by Keilin in the early 1900s (Keilin, 1921) and was placed in a separate order, Helicosporidia, within Cnidiospora (Protozoa) by Kudo (1931). Since then, additional helicosporidians have been detected in mites, cladocerans, trematodes, collembolans, scarabs, mosquitoes, simuliids, and pond water samples (Kellen and Lindegren, 1973; Fukuda et al., 1976; Sayre and Clark, 1978; Purrini 1984; Avery and Undeen, 1987). Weiser (1964, 1970) examined the type material and a new isolate of Helicosporidia from a hepialid larva, and he proposed that this organism should be transferred to the Ascomycetes, because of some analogies in pathways of infection. Additionally, Kellen and Lindegren (1974) isolated a *Helicosporidium* from infected larvae and adults of *Carpophilus mutilatus* (Coleoptera: Nitidulidae) and described its life cycle in a lepidopteran host, the navel orangeworm *Paramyelois transitella*. They agreed that this organism is not a protozoan but remained uncertain about its taxonomic position. Later, Lindegren and Hoffman (1976) proposed that the developmental stages of this organism placed it closer to the

Protozoa than to the Fungi. Because of this uncertain taxonomic status, the Helicosporidia have not appeared in classification systems of either the Protozoa or the Fungi (Cavalier-Smith, 1998; Tehler et al., 2000).

Recently, a *Helicosporidium* sp. isolated from the blackfly *Simulium jonesi* Stone and Snoddy (Diptera: Simuliidae) has been shown to replicate in a heterologous host *Helicoverpa zea* (Lepidoptera: Noctuidae), which has provided a means to produce quantities sufficient for density gradient extraction of the infectious cyst stage (Boucias et al., 2001). In order to evaluate the taxonomic position of this *Helicosporidium* sp. within the eukaryotic tree, we extracted genomic DNA from the cyst preparation and PCR-amplified several targeted genes (5.8S, 28S, 18S ribosomal regions, partial sequences of the actin and β-tubulin genes). These genes were selected because they have been used extensively to infer deep eukaryotic phylogenies (Philippe and Adoutte, 1998). Amplified genes were sequenced and information from nucleotide sequences was subjected to comparative analysis.

Materials and Methods

Cyst Preparation and DNA Extraction

Helicosporidium sp. was originally isolated from the blackfly *Simulium jonesi* Stone and Snoddy (Diptera: Simuliidae) and produced in *Helicoverpa zea* (Boucias et al., 2001). Approximately $4x10^7$ cysts suspended in 0.15 M NaCl were applied to a linear gradient of 1.00-1.3003 g ml^{-1} of Ludox HS40 (DuPont). Helicosporidial cysts that banded at an estimated density of 1.17 g ml^{-1} were collected, diluted in ten volumes of deionized H_2O, and washed free of residual Ludox by repeated low-speed centrifugation steps. The pellet, resuspended in 50 µl of H_2O, was extracted with the use of the

Masterpure^{Tm} Yeast DNA extraction kit (Epicentre Technologies), following the manufacturer's protocol. Examination of the cells before and after lysis treatment revealed the presence of numerous, highly refractile cysts before treatment, and, after incubation in the lysis buffer at 50 °C, cysts appeared to dehisce, releasing the filamentous cells. However, no massive disruption of the ovoid cells or filamentous cells was observed in these preparations. Visible pellets were observed after RNase treatment, phenol-chloroform extraction, and ethanol precipitation. The final pellet, suspended in molecular biology grade water, was frozen at -20 °C.

Amplification, Cloning and Sequencing of Extracted DNA

The ITS1-5.8S-ITS2, 28S, and 18S ribosomal regions of the helicosporidial DNA were amplified with a mixture of *Taq* DNA polymerase (Promega) and *PFU* polymerase (Stratagene), using the primers TW81 and AB28 for the ITS-5.8S (Curran et al., 1994) and NL-1 and NL-4 primers for the 28S (Kurtzman and Robnett, 1997). Two primer sets (sequences in Appendix A) designed from consensus regions of selected protist sequences downloaded from GenBank were used to amplify the 18S region. Several series of primers, also designed from consensus regions of selected protist genes, were used to PCR-amplify partial sequences of the actin and β-tubulin genes. All primer sequences are listed in Appendix A. DNA was excised from agarose gels, extracted with the QiaxII gel extraction kit (Qiagen), and sent to the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida for direct sequencing.

DNA Sequence Analysis

The helicosporidial 18S region sequence was aligned with 138 other sequences from representative eukaryotic taxa obtained from the Ribosomal Database Project (RDP,

Maidak et al., 2000). Downloaded sequences were pre-aligned based on the secondary structure of the rDNA. An additional 18S sequence from the pathogenic alga *Prototheca wickerhamii* was downloaded from GenBank (accession number X56099) and incorporated in the SSU-RNA data set. Additionally, eukaryotic 28S sequences were downloaded from GenBank and aligned with the helicosporidial 28S sequence using ClustalX (Thompson et al., 1997). Eventually, SSU- and LSU-rDNA data sets were combined to infer one single ribosomal phylogeny. Both *Helicosporidium* sp. actin and β-tubulin nucleotide sequences were aligned with homologous sequences downloaded from GenBank. Alignments were obtained using ClustalX software with default parameters. All data sets were checked by eye before further analyses in order to insure that no region of uncertain alignment was present. The final aligned data sets can be obtained from TreeBase (Morel, 1996; http://www.herbaria.harvard.edu/treebase) with the study accession number S604. The 18S algal alignment was kindly provided by V. A. R. Huss, from the University of Erlangen, Germany.

Aligned data sets were subjected to a partition homogeneity test using the program PAUP*, version 4.0b4a (Swofford, 2000), in order to assess the extent of character incongruence between the data sets (Farris et al., 1994). Phylogenies were then reconstructed using Neighbor-Joining (NJ) as implemented in the PAUP* program version 4.0b4a. Neighbor-Joining analyses were based on the Paralinear/LogDet model of nucleotide substitution (Lockhart et al., 1994). This method allows for nonstationary changes in base composition and has been shown to reduce support for spurious resolutions, such as Long Branch Attraction (Felsenstein, 1978). Monophyly of groups was assessed with the bootstrap method (100 replicates). Additionally, maximum-

parsimony analyses, including jackknifing (100,000 replicates, Farris et al., 1996) were also performed using PAUP*. We chose the latter, conservative approach for its ability to rapidly search a large amount of tree space and estimate support for unambiguously resolved groups (Lipscomb et al., 1998).

Results

Five PCR-amplified gene fragments of the *Helicosporidium* sp. were sequenced. These sequences corresponded to the 18S, 28S, ITS1-5.8S-ITS2, actin and β-tubulin genes, and were 1558, 661, 844, 880 and 879 bases in length, respectively. The DNA nucleotide sequences have been submitted to the GenBank database with respective accession numbers: AF317893, AF317894, AF317895, AF317896 and AF317897. All sequences, examined by BLAST analysis (Altschul et al., 1997), produced matches with extremely low Expect (E) values. Two algal species, *Chlamydomonas reinhardtii* and *Volvox carteri*, were highly similar to all five sequences. Additionally, other algal genera, such as *Trebouxia*, *Scenedesmus*, or *Chlorella*, were found to match recurrently with the helicosporidial sequences.

A preliminary partition homogeneity test showed that the 18S, 28S and 5.8S sequences were highly concordant (data not shown). A first phylogenetic tree was inferred from the 18S sequence aligned with the 140 sequences downloaded from the RDP website. This tree placed *Helicosporidium* sp. as a member of the green algae, and this association was supported by significant bootstrap values (data not shown). The tree presented in Fig. 2-1 was inferred from a combined data set SSU+LSU rDNA, and is concordant with the preliminary result. This tree was rooted by using *Dictyostelium discoideum* as an outgroup (Fig. 2-1). Although the taxonomic position of *D. discoideum*

is subject to debate (Baldauf et al., 2000), it appears basal in conservative rDNA reconstruction (Lipscomb et al, 1998). Our tree is fairly consistent with other previous molecular phylogenetic studies of eukaryotes (Drouin et al., 1995, Lipscomb et al., 1998, Baldauf et al., 2000), showing that the animal and fungal lineages share a more recent common ancestor than either does with the plant lineage (Baldauf and Palmer, 1993) and that green algae and green plants form a monophyletic group (Fig. 2-1). Due in part to limited sampling, the relationships between protists are not well resolved, but they all appear near the root of the tree (Fig. 2-1). Importantly, the tree shows that *Helicosporidium* sp. clusters with the green algae (Chlorophyta), and this relationship is supported by both Neighbor-Joining (89) and maximum parsimony (69) bootstrap/jackknife methods (Fig. 2-1).

The tree presented in Fig. 2-2 was inferred from an algal SSU-rDNA alignment, and it addresses the position of *Helicosporidium* sp. within the Chlorophyta. This tree is rooted with the branch leading to Charophyte algae and shows the four classes of Chlorophyta. As previously shown by Bhattacharya and Medlin (1998), the class Prasinophyceae is paraphyletic, whereas Ulvophyceae, Trebouxiophyceae, and Chlorophyceae are monophyletic. In this tree, *Helicosporidium* sp. is depicted as a sister taxon to *Prototheca zopfii* (Trebouxiophyceae) by both distance and parsimony analyses (Fig. 2-2).

Preliminary alignments showed that both actin and β-tubulin genes amplified from helicosporidial DNA did not possess any introns. As a result, these sequences were aligned with homologous coding sequences (cDNA) downloaded from GenBank. The phylogenetic trees inferred from the analysis of actin and β-tubulin fragments are

presented in Figs. 2-3 and 2-4, respectively. Both trees are very similar: they are rooted with the branch leading to the ciliate *Euplotes crassus*, and they present branching patterns common to most eukaryotic phylogenies. All protists are clustered near the root of the trees, and Metazoa, Fungi, and Viridiplantae all are shown to be monophyletic. Both trees confirm that *Helicosporidium* sp. belongs to the green algae clade, even if the resolution within this clade is not very high (Fig. 2-3 and 2-4). Once again, the nodes linking *Helicosporidium* sp. to green algae are all supported, except for the parsimony jackknife of the β-tubulin tree (Fig. 2-4).

Additionally, further analyses led to the same conclusion that *Helicosporidium* sp. groups with the green algae. Notably, realignments of the RDP SSU-rDNA data set, modification of gap penalty parameters or utilization of other distance methods available in PAUP* (such as HKY85 or maximum likelihood distance) had no effect on the final position of *Helicosporidium* sp. within the eukaryotic tree.

Discussion

All trees obtained in this phylogenetic study present a reasonable branching pattern, with major divisions corresponding to conventional taxonomic classification (Kinetoplastida, Alveolata, Viridiplantae, Fungi and Metazoa). On the basis of these phylogenies, *Helicosporidium* sp. is unrelated to any group of Protozoa (Philippe and Adoutte, 1998). This result suggests that Kudo's early attempt (1931) to classify this organism within the Protozoa may have been wrong, but it is consistent with studies by Weiser (1970) and by Kellen and Lindegren (1974), who both proposed the removal of the Helicosporidia from the Protozoa. However, in a more recent study, Lindegren and Hoffman (1976) refused this suggestion and re-affirmed that the Helicosporidia have

affinities with the Protozoa, based on the presence of well-defined Golgi bodies and mitotic division of the nucleus.

None of the phylogenic trees depicted *Helicosporidium* sp. as a member of the kingdom Protozoa (as defined by Cavalier-Smith, 1993). Instead, they consistently and stably grouped *Helicosporidium* sp. among members of Chlorophyta, suggesting that this invertebrate pathogen is a green alga. Considering the fact that comparative sequence analysis is a robust method that provides resolving power for clade identification, the appropriate place of *Helicosporidium* is within the Chlorophyta. Furthermore, the 18S-based phylogeny of the Chlorophyta depicted *Helicosporidium* sp. as a member of the class Trebouxiophyceae and as a very close relative to the genus *Prototheca* (Fig. 2-2). In these 18S trees, *Helicosporidium* sp. always appears as sister taxon to *P. zopfii*, and the relationship is always supported by bootstrap and jackknife analyses.

It may be argued that the helicosporidial sequences, because they were amplified with universal primers, may have resulted from a potential algal contaminant. However, it should be noted that our *Helicosporidium* sp. was carefully purified by gradient centrifugation after propagation in *Helicoverpa zea*. Furthermore, Boucias et al. (2001) also propagated *Helicosporidium* sp. *in vitro* and extracted DNA from both *in vitro* and *in vivo* sources. An RFLP analysis of the 18S gene amplified from these two sources produced identical digest patterns, demonstrating the integrity of the extracted helicosporidial genomic DNA used in this study (Boucias et al., 2001). Also, DNA has been extracted from a second strain of *Helicosporidium* sp., and SSU-rDNA gene sequences from both strains are highly similar (see Appendix B).

The association of *Helicosporidium* sp. with the genus *Prototheca* is interesting from a biological perspective. Members of both genera are achlorophylous and are animal pathogens. To date, *Helicosporidium* spp. have been identified as invertebrate pathogens, whereas *Prototheca* spp. are known to be pathogenic to vertebrates, including humans (Galan et al., 1997; Mohabeer et al., 1997). Mohabeer et al. (1997) reported that *Prototheca wickerhamii*, although being primarily infectious to the skin, can invade several human tissues, including the liver, spleen, small intestine, lymph nodes, central nervous system, and blood. *Prototheca zopfii* is also reported to be a human pathogen (Galan et al., 1997). Morphologically, the vegetative cells of the *Helicosporidium* sp. produced under *in vitro* and *in vivo* conditions are reminiscent of that reported for the genus *Prototheca*. Indeed, as protothecans, the vegetative cells of *Helicosporidium* sp. undergo one or two cell divisions within a pellicle. This pellicle eventually splits open or dehisces, releasing either two or four daughter cells from the parent cell wall or pellicle (Boucias et al., 2001). However, protothecans have yet to be reported to produce a mature cyst containing the filamentous cell, which is the very unique morphological feature that characterizes the genus *Helicosporidium*. Deeper analyses, as well as cell biology observations (Taylor, 1999), will likely confirm the relationship between the genera *Helicosporidium* and *Prototheca*. Notably, comparative analysis of mitochondrial genomes has been shown to be a very powerful tool for classification of green algae (Nedelcu et al., 2000).

Both morphological and molecular evidence suggest that the appropriate place of the group Helicosporidia is within the green algae. Therefore, the genus *Helicosporidium*

represents the first reported algal entomopathogen, and it should be placed among the Chlorophyta, Trebouxiophyceae.

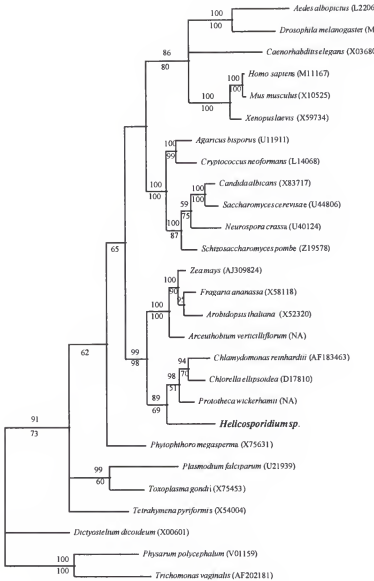


Figure 2-1: Phylogram inferred from combined SSU-rDNA and LSU-rDNA nucleotide sequence alignment, showing that *Helicosporidium* sp. is grouped with green algae. Numbers at the top of the nodes represent the results of bootstrap analyses (100 replicates) using Neighbor-Joining method. Numbers at the bottom of the nodes are results of parsimony jackknife analyses (100,000 replicates). Only values superior to 50% are shown. SSU-rDNA sequences were downloaded from the Ribosomal Database Project (RDP) website. LSU-rDNA sequences were downloaded from GenBank. Accession numbers for these sequences are indicated after each species name (NA: LSU sequence not available in GenBank).

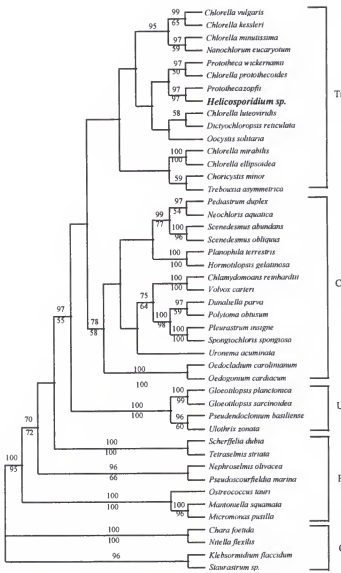


Figure 2-2: SSU-rDNA phylogeny of Chlorophyte green algae. *Helicosporidium* sp. appears as a member of the class Trebouxiophyceae, sister taxon to *P. zopfii*. Numbers at the top of the nodes represent the results of bootstrap analyses (100 replicates) using Neighbor-Joining method. Numbers at the bottom of the nodes are results of jackknife analyses (100,000 replicates) using Maximum-Parsimony method. Only values superior to 50% are shown. The tree is rooted with Charophyte green algae.

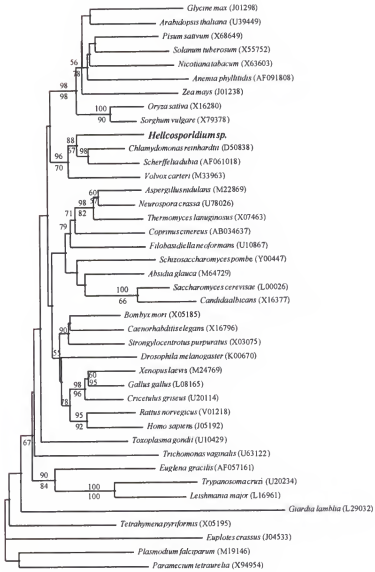


Figure 2-3: Phylogenetic tree based on actin gene nucleotide sequences. The tree depicts *Helicosporidium* sp. as a Chlorophyta. Numbers at the top of the nodes represent the results of bootstrap analyses (100 replicates) using Neighbor-Joining method. Numbers at the bottom of the nodes are results of jackknife analyses (100,000 replicates) using Maximum-Parsimony method. Only values superior to 50% are shown. All but the helicosporidial sequences were downloaded from GenBank. Accession numbers for these sequences are indicated after each species name.

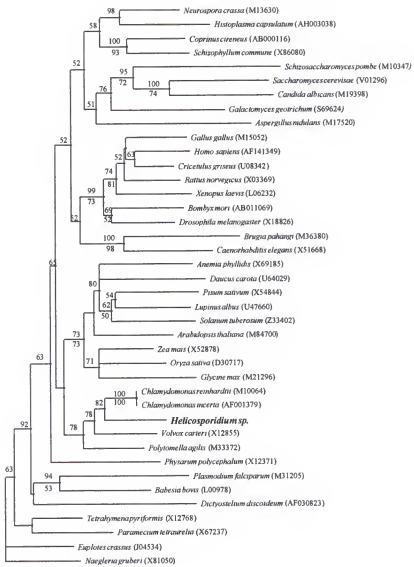


Figure 2-4: Phylogenetic tree based on β-tubulin gene nucleotide sequences. In this tree, *Helicosporidium* sp. appears as sister taxa to the genus *Chlamydomonas*. Numbers at the top of the nodes represent the results of bootstrap analyses (100 replicates) using Neighbor-Joining method. Numbers at the bottom of the nodes are results of jackknife analyses (100,000 replicates) using Maximum-Parsimony method. Only values superior to 50% are shown. All but the helicosporidial sequences were downloaded from GenBank. Accession numbers for these sequences are indicated after each species name.

CHAPTER 3
ORGANELLAR GENE PHYLOGENIES

Introduction

The Helicosporidia have been detected in insects, collembolans, mites, crustaceans, and trematodes, and they also have been isolated from ditch water samples (Kellen and Lindegren, 1973; Sayre and Clark, 1978; Purrini, 1984; Avery and Undeen, 1987a; Pekkarinen, 1993). These pathogens have a worldwide geographical range and have been found in Europe, South America, North America, Asia, and Africa (Keilin, 1921; Weiser, 1970; Kellen and Lindegren, 1973; Hembree, 1979; Seif and Rifaat, 2001). Although *Helicosporidium* spp. seem to be ubiquitous, they have been studied so little that their occurrence and their importance as invertebrate pathogens are unclear. Recently, a *Helicosporidium* sp. was isolated from larvae of the black fly *Simulium jonesi* Stone and Snoddy collected in Florida (Boucias et al., 2001). Microscopic observation of the vegetative growth of *Helicosporidium* sp. under *in vivo* and *in vitro* conditions led Boucias et al. (2001) to associate this protist with green algae, particularly the unicellular, non-photosynthetic, and pathogenic algae belonging to the genus *Prototheca*. Boucias et al. (2001) noticed that, as protothecans, the vegetative cells of *Helicosporidium* sp. undergo one or two cell divisions within a pellicle. This pellicle eventually splits open and releases either two or four daughter cells. This association between *Helicosporidium* and *Prototheca* was surprising but was later confirmed by molecular sequence comparisons (see Chapter 2). Phylogenetic analyses of several *Helicosporidium* sp. genes (rDNA, actin and β-tubulin) all identified this organism as a member of the green algae

clade (Chlorophyta). Moreover, a nuclear 18S rDNA phylogeny of the Chlorophyta depicted *Helicosporidium* sp. as a close relative of both *Prototheca wickerhamii* and *Prototheca zopfii* within the class Trebouxiophyceae. Based on both morphological and molecular evidence, the transfer of the genus *Helicosporidium* to Chlorophyta, Trebouxiophyceae was proposed.

Prototheca spp. have been shown to be closely related to the photoautotrophic genus *Chlorella* (Chlorophyta, Trebouxiophyceae), based on phylogenetic analyses inferred from the nuclear 18S rDNA and the plastid 16S rDNA genes (Huss et al., 1999; Nedelcu, 2001). The plastid 16S rDNA gene (*rrn16*) is a chloroplast gene. Despite having lost their photosynthetic abilities, non-photosynthetic green algae such as protothecans have been found to retain vestigial, degenerate chloroplasts called leucoplasts. The presence of such plastids has been demonstrated extensively in the non-photosynthetic green algae of the genus *Polytoma* (Lang, 1963; Siu et al., 1976), which are closely related to *Chlamydomonas* spp. (Chlorophyta, Chlorophyceae). In contrast, there are no records of microscopic observations of a leucoplast in a *Prototheca* sp. cell. However, the plastid genome of *Prototheca wickerhamii* recently has been isolated and partially sequenced (Knauf and Hachtel, 2002). Similar to the situation described previously for plastid genomes in non-photosynthetic plants (reviewed in Hachtel, 1996), this genome is highly reduced in size but is believed to be functional.

In addition, *P. wickerhamii* also is known to possess a very characteristic mitochondrial genome. As reviewed by Nedelcu et al. (2000), the *Prototheca*-like mitochondrial genome represents an ancestral type among green algae that features

(among other characteristics) a larger size (45-55 kb) and a more complex set of protein-coding genes than the derived, *Chlamydomonas*- mitochondrial genome.

In order to confirm *Helicosporidium* sp. as a green alga and as a close relative of the genus *Prototheca*, the presence of organellar (mitochondrial and plastid) DNA in helicosporidial cells was investigated. This chapter reports the PCR amplification and sequencing of mitochondrial *cox3* and plastid *rrn16* homologues from *Helicosporidium* sp. Moreover, these genes were also used to infer organellar gene-based phylogenies of the Chlorophyta that includes the genus *Helicosporidium*.

Materials and Methods

Helicosporidium Isolate

The *Helicosporidium* sp. was isolated from the black fly *Simulium jonesii* and was successfully amplified in *Helicoverpa zea* larvae, as previously described (Boucias et al., 2001). Cysts produced in *H. zea* larvae were purified by gradient centrifugation on Ludox and grown in artificial media (TNM-FH insect medium, supplemented with gentamicin and 5% fetal bovine serum, Sigma-Aldrich) before harvest and DNA extraction.

DNA Extraction and Amplification

Helicosporidial DNA was extracted according to Boucias et al. (2001) using the Masterpure Yeast DNA extraction kit from Epicentre Technologies. Cellular DNA was used as a template for the PCR amplification of the *rrn16* gene using chloroplast 16S rDNA gene specific primers ms-5' and ms-3' listed by Nedelcu (2001). The helicosporidial cox3 homologue was amplified using the primers CC66 and CC67 (see Appendix A for primer sequences). PCR products were gel-purified with the QiaxII gel extraction kit (Qiagen) and cloned in pGEM-T vectors using the pGEM-T easy vector

systems (Promega). Positive clones were sent to the Interdisciplinary Core for Biotechnology Research (ICBR) at the University of Florida for sequencing.

Phylogenetic Analyses of the *rrn16* Sequence

The plastid 16S rDNA sequence from *Helicosporidium* sp. was aligned with homologous sequences available in GenBank. The alignment was obtained using ClustalX software with default parameters (Thompson et al., 1997) and optimized manually. Analyses of the aligned sequences were performed in PAUP* version 4.0 beta 10 (Swofford, 2000), using maximum parsimony (MP) and neighbor joining (NJ) methods. MP analyses were performed using the default parameters in PAUP*. NJ analyses were based on the two-parameter method of Kimura, but other models, including HK85 and the three-parameter Kimura method, were also used. Branch support for MP and NJ analyses was assessed by bootstrapping (100 replicates). The alignment, as well as the resulting trees, can be obtained from TreeBase (Morell, 1996; http://www.treebase.org), with the study accession number S819.

Phylogenetic Analyses of the *cox3* Sequence

The *cox3* gene from *Helicosporidium* sp. was translated *in silico*, and the resulting amino acid sequence was then aligned with homologous protein fragments downloaded from GenBank (using the ClustalX algorithm). Phylogenetic relationships were inferred using the NJ and MP algorithms in PAUP*. Bootstrap support was calculated for both methods (100 replicates).

Results

Amplification of *Helicosporidium* sp. Organellar Genes

Fragments homologous to mitochondrial *cox3* and plastid *rrn16* genes were successfully amplified from the *Helicosporidium* cellular DNA preparation. The fragment

lengths are 412 bp for the *Helicosporidium cox3* gene and 1266 bp for the *Helicosporidium rrn16* gene. Both sequences are available in the GenBank public database with the accession numbers AY445515 and AF538864 for the *cox3* and *rrn16* genes, respectively. The two gene sequences are very similar to homologous genes previously sequenced from other green algae. Both genes are very AT-rich: 60.7% for the *rrn16* sequence and 65.8% for the *cox3* gene. Such a deviation from homogeneity is common in nonphotosynthetic algal genes; for example, the AT content of the *Prototheca zopfii* plastid 16S rDNA gene is 63.1% (Nedelcu, 2001). Similarly, the mitochondrial *cox3* gene of *P. wickerhamii* has also been found to be very AT-rich (66.7%; Wolff et al., 1994).

Phylogenetic Analyses

The plastid 16S rDNA gene sequence was compared with 21 homologous sequences from algal species belonging in two major classes of Chlorophyta - Trebouxiophyceae and Chlorophyceae. Both classes include some non-photosynthetic species. Phylogenetic reconstructions using Neighbor-Joining and Parsimony methods produced the same tree, presented in Fig. 3-1. The MP/NJ tree (Fig. 3-1) was rooted with the plastid 16S rDNA sequence of *Nephroselmis olivacea*, a member of the class Prasinophyceae, which is thought to include descendants of the earliest-diverging green algae (Turmel et al., 1999). The relationships among green algal taxa depicted in Fig. 3-1 are consistent with affiliations previously suggested by other phylogenetic studies (Bhattacharya and Medlin, 1998; Huss et al., 1999; Nedelcu, 2001; see also Chapter 2). First, both classes (Trebouxiophyceae and Chlorophyceae) appear monophyletic. Within the Chlorophyceae, two nonphotosynthetic clades can be identified (Fig. 3-1); *Polytoma*

uvella, *P. obtusum* and *P. mirum* are monophyletic and are sister taxa to *Chlamydomonas applanata*, whereas *P. oviforme* is more closely related to *C. moewusii*. A paraphyletic *Polytoma* has previously been demonstrated by Nedelcu (2001) based on nuclear 18S rDNA and plastid 16S rDNA phylogenies. Only one non-photosynthetic clade exists among the Trebouxiophyceae (as identified by Nedelcu, 2001). This clade is strongly supported by bootstrap values, and it includes *Helicosporidium* sp., *Prototheca* spp., and *Chlorella protothecoides*, an auxotrophic, mesotrophic, but photosynthetic species. The genus *Prototheca* appears paraphyletic, as previously shown by nuclear 18S rDNA and plastid 16S rDNA phylogenies (Huss et al., 1999; Nedelcu, 2001). In the tree (Fig. 3-1), *Helicosporidium* sp. is depicted as being a sister taxon to *Prototheca zopfii*, and this relationship is supported by maximal bootstrap values. This is consistent with previous nuclear 18S rDNA phylogenies (Chapter 2).

The *cox3* fragment amplified from *Helicosporidium* sp. DNA is also very similar to green algal homologous genes. However, compared to the rrn16 gene, fewer *cox3* homologous sequences are available publicly. The helicosporidial *cox3* fragment translation was aligned with 5 other sequences, and the phylogenetic tree inferred from this alignment is presented in Fig. 3-2. As it is the case for the *rrn16* phylogenies, both NJ and MP methods led to the same tree topology, and the *Nephroselmis olivacea* homologue was used to root the trees. The tree identifies two monophyletic clades that correspond to two Chlorophyta classes: Trebouxiophyceae and Chlorophyceae. Confirming the results previously obtained in other phylogenies, the tree depicts *Helicosporidium* sp. as a sister taxon to *Prototheca wickerhamii*, within the class

Trebouxiophyceae. This relationship, once again, is supported strongly by bootstrapping, in both parsimony and distance trees (Fig. 3-2).

Discussion

Presence of Organelle-Like Genes and Genomes

The presence of mitochondrial and plastid genes strongly suggests that *Helicosporidium* cells may contain such organelles and their respective genomes. By itself, the existence of such organelles provides additional evidence for the taxonomic classification of the Helicosporidia. For example, the fact that *Helicosporidium* sp. seems to contain mitochondria suggests that the Helicosporidia are not related to the amitochondriate Microsporidia (as was proposed by Kudo, 1931). Although some mitochondrial-like genes have been amplified from microsporidian DNA preparation (Keeling and Fast, 2002), only a few genes are involved, and cox3 has not been one of them. More importantly, the presence of chloroplasts, even if they are probably highly reduced, provides strong arguments in favor of Helicosporidia being non-photosynthetic green algae. However, this evidence is not sufficient to affirm that *Helicosporidium* sp. belongs to the Chlorophyta. Indeed, other protists, most notably the phylum Apicomplexa, have also been shown to possess a degenerate, vestigal chloroplast (apicoplast) with a functional genome (Wilson, 2002). This plastid has been proposed to derive from an endosymbiotic interaction with a red alga (secondary symbiosis). The algal nature of *Helicosporidium* already has been suggested by morphological observations (Boucias et al., 2001) and strongly supported by phylogenetic analyses inferred from several nuclear genes (Chapter 2). Therefore, helicosporidial cells are likely to possess a plastid similar to other non-photosynthetic Chlorophyta, derived from a primary endosymbiosis.

In contrast to the nuclear genome, where only a few genes have been sequenced, there is much information on both *Prototheca wickerhamii* mitochondrial and plastid genome sequences (Wolff et al., 1994; Knauf and Hachtel, 2002). Therefore, the sequencing of *Helicosporidium* sp. organellar genes also provides an opportunity for more sequence comparison analyses.

Phylogenetic Analyses

Comparative analyses of the mitochondrial and plastid gene sequences confirm that Helicosporidia are closely related to non-photosynthetic algae in the class Trebouxiophyceae (Chlorophyta). The *rrn16* phylogenies are much more robust, because they include many more species. In all *rrn16* phylogenetic trees, *Helicosporidium* sp. appears as member of the *Prototheca* clade (as defined by Nedelcu, 2001), sister taxon to *Prototheca zopfii*. The position of *Helicosporidium* spp. is identical in phylogenies based on nuclear 18S rDNA genes (Chapter 2). Similar to the situation observed in the 18S rDNA phylogeny, the branch leading to the *Helicosporidium* + *P. zopfii* clade is the longest of the tree, suggesting that this association could be an artifact due to long-branch attraction. However, it should be noted that *Helicosporidium* spp. are depicted in exactly the same position even if *P. zopfii* is removed from the sequence alignment, and their relationship with *P. wickerhamii* is still very strongly supported (data not shown). Therefore, this relationship is not an artifact.

Based on all of these phylogenetic analyses (Chapters 2 and 3), the Helicosporidia should be included in the *Prototheca* clade defined by Nedelcu (2001). The clade is consistently and strongly supported by resampling tests, suggesting that *Helicosporidium* sp., *Prototheca* spp., and *Chlorella protothecoides* may have arisen from a common

ancestor. Within the clade, the relationships are less robust; the genus *Prototheca* has always appeared paraphyletic, and *Chlorella protothecoides*, despite being proposed to be the closest green relative of *Prototheca* spp., has never appeared in a basal position (Huss et al., 1999; Nedelcu, 2001; see also Chapter 2). In the more complete *rrn16* trees (Fig. 3-1), these ambiguities remain. However, additional resolution may be obtained inside the *Prototheca* clade by adding more taxa and/or by using other genes, such as protein-encoding genes, which are likely to exhibit a lower rate of nucleotide substitution.

The *Helicosporidium* sp. *cox3* gene encodes for a protein (cytochrome *c* oxidase subunit 3) and exhibits a lower rate of substitution, as shown by the length of the branch leading to *Helicosporidium* sp. in phylogenetic trees (Fig. 3-2). However, *cox3*-inferred phylogenies do not allow for extensive comparison because there are too few homologous sequences within the green algae. They do provide confirmation that *Helicosporidium* and *Prototheca* are closely related genera.

Prototheca-Like Organelle Genomes

Phylogenetic affinities and the presence of two organellar genes (mitochondrial *cox3* and plastid *rrn16*) suggest that the Helicosporidia possess a mitochondrial genome and a plastid genome similar to *P. wickerhamii*. In this non-photosynthetic alga, the size of the chloroplast (leucoplast) genome has been estimated to be 54,100 bp, which is much smaller than the 150 kb chloroplast DNA of the photosynthetic relative *Chlorella vulgaris* (Knauf and Hachtel, 2002). This decrease in size is common in all secondary, non-photosynthetic green plants and algae (Hachtel, 1996) and has been explained by the loss of most of the plastid genes that were involved in photosynthesis. However, some plastid genes have been selectively retained, suggesting that they may encode for

essential protein products. In *Prototheca*, the functions of these proteins are not known (Knauf and Hachtel, 2002). In Apicomplexa, retained plastid ORFs have been associated with the apicoplast's hypothetical primary functions: fatty acid and isoprenoid biosynthesis (reviewed by Wilson, 2002).

Additionally, *P. wickerhamii* also is known to possess a characteristic mitochondrial genome within the green algae. This genome has been entirely sequenced (Wolff et al., 1994), and it has subsequently been shown to be significantly different from other algal genomes. The *Prototheca*-like mitochondrial genome represents an ancestral type among green algae, as opposed to the more derived *Chlamydomonas*-like mitochondrial genome (reviewed by Nedelcu et al., 2000). One major difference between the two types of algal mitochondrial genomes is the presence or absence of the *cox3* gene. In the green alga *Chlamydomonas reinhardtii* and the colorless alga *Polytomella* sp., the *cox3* gene has been transferred from the mitochondrial genome to the nucleus (Perez-Martinez et al., 2000). In *Prototheca wickerhamii*, the *cox3* gene has been conserved in the mitochondrial genome (Wolff et al., 1994). The Chlorophyceae *Scenedesmus obliquus* presents an intermediate type of algal mitochondrial genome that includes the *cox3* gene (Nedelcu et al., 2000). According to the sequence comparison analysis, it is likely that the *Helicosporidium* sp. *cox3* homologue is present in the helicosporidial mitochondrial genome.

Having shown that the Helicosporidia are non-photosynthetic green algae and close relatives to the genus *Prototheca*, a logical hypothesis is that *Helicosporidium* sp. possesses *P. wickerhamii*-like organelles and organelle genomes, i.e., a highly reduced plastid genome and an ancestral type of mitochondrial genome.

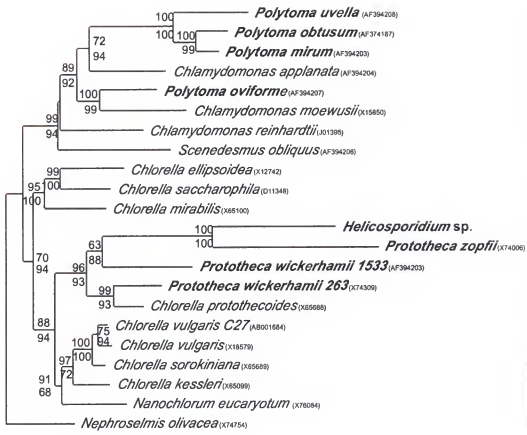


Figure 3-1: Phylogenetic tree based on plastid 16S rDNA sequence. *Helicosporidium* sp. is depicted as Trebouxiophyceae, member of a strongly supported *Prototheca* clade, and sister taxa to *Prototheca zopfii*. Non-photosynthetic taxa are in bold. Branch lengths correspond to evolutionary distances. Numbers at the top and bottom of the nodes represent the results of bootstrap analyses (100 replicates) using Maximum-Parsimony and Neighbor-Joining methods, respectively. Only values greater than 50% are shown. All but the helicosporidial sequences were downloaded from GenBank. Accession numbers for these sequences are indicated after each species name.

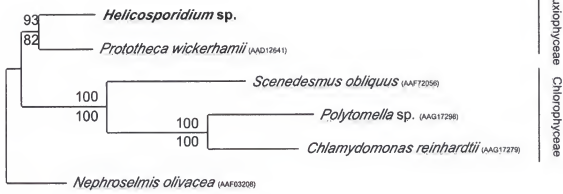


Figure 3-2: Phylogram inferred from a *cox3* gene fragment alignment. The tree depicts *Helicosporidium* sp. as a Trebouxiophyceae, sister taxa *to Prototheca wickerhamii*. Branch lengths correspond to evolutionary distances. Numbers at the top and bottom of the nodes represent the results of bootstrap analyses (100 replicates) using Maximum-Parsimony and Neighbor-Joining methods, respectively. Only values greater than 50% are shown. All but the helicosporidial sequences were downloaded from GenBank. Accession numbers for these sequences are indicated after each species name.

CHAPTER 4
INVESTIGATION ON THE *HELICOSPORIDIUM* SP. PLASTID GENOME

Introduction

The Helicosporidia are obscure pathogenic protists that have been reported in a wide range of invertebrate hosts (Keilin, 1921; Weiser, 1970; Kellen and Lindegren, 1973; Fukuda et al., 1976; Sayre and Clarke, 1978; Hembree, 1979; Purrini, 1984; Pekkarinen, 1993; Seif and Rifaat, 2001). They are characterized by the formation of a highly resistant cyst that encloses three ovoid cells and a diagnostic filamentous cell (Keilin, 1921). To date, it remains unclear whether the Helicosporidia possess a free-living stage or are obligate pathogens that exist outside their hosts only as cysts.

A new *Helicosporidium* sp. was recently isolated in Florida (Boucias et al., 2001). Morphological and molecular data compiled on this organism have demonstrated that the Helicosporidia are non-photosynthetic green algae, and they are related to *Prototheca*, another non-photosynthetic, parasitic algal genus (Boucias et al., 2001; Chapters 2 and 3; see also Ueno et al., 2003). Furthermore, sequencing of chloroplast-like molecules has provided evidence that both *Prototheca* and *Helicosporidium* have retained a modified chloroplast and chloroplast genome (Chapter 3; Knauf and Hachel, 2002). The presence of plastid-like structures in *Prototheca zopfii* has also been suggested following microscopic observations (Melville et al., 2002).

Cryptic, modified chloroplasts (and their genomes) have been reported in a variety of non-photosynthetic protists, including the green algae *Prototheca wickerhamii* (Knauf and Hachel, 2002), the euglenoid *Astasia longa* (Gockel and Hachtel, 2000), the

stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Sekigushi et al., 2002) and the apicomplexan parasites *Plasmodium falciparum* and *Toxoplasma gondii* (reviewed by Wilson, 2002). Sequence information on secondary, non-photosynthetic plastid genomes is accumulating, showing that these genomes are much smaller than that of photosynthetic relatives, but they have remained functional. A widely accepted hypothesis is that the reduction in size can be explained by the loss of most of the genes involved in photosynthesis. The remaining genes have been selectively retained because they are involved in other essential plastid function(s). Whether all the secondary non-photosynthetic plastids have been retained for the same reasons is unclear, as the number of retained plastid genes varies depending on the species. As reviewed by Williams and Keeling (2003), the plastid genomes of parasitic organisms (*Plasmodium falciparum*, *Prototheca wickerhamii*) tend to be more reduced.

The *Helicosporidium* sp. plastid genome is expected to be similar to that of *Prototheca wickerhamii* (estimated at 54 kb; Knauf and Hachtel, 2002). In an effort to better characterize the *Helicosporidium* sp. vestigial chloroplast, a portion of the plastid genome has been sequenced and compared to two close relatives: the *Prototheca wickerhamii* plastid genome (Knauf and Hachel, 2002) and the *Chlorella vulgaris* chloroplast genome (Wakasugi et al., 1997).

Materials and Methods

Helicosporidium Isolate and Culture Conditions

The *Helicosporidium* sp. was originally isolated from a black fly larvae (Boucias et al., 2001). It was maintained *in vitro* in Sabouraud Maltose agar supplemented with 2% Yeast extract (SMY) at 25°C. Helicosporidial cells produced on these plates were inoculated into flasks containing SMY broth and shaken at 23°C on a rotary shaker (250

rpm) for 3-4 days. Cells were collected by centrifugation and used for DNA extraction. In addition, helicosporidial cysts were collected from laboratory-infected *Helicoverpa zea*, purified by Ludox gradient centrifugation, and stored in sterile water at 4°C, following a protocol previously described by Boucias et al. (2001).

CHEF Gel Electrophoresis

Helicosporidial cysts (ca. 1.5 x 10^8 cysts) were incubated in DMSO (100%) at room temperature for 30 minutes. They were then collected by centrifugation and resuspended in 200 µl of 10 mM TrisHCl, 50 mM EDTA buffer. After mixing quickly with 200 µl of 2% low-melting-point agarose in 10 mM TrisHCl, 50 mM EDTA buffer, the *Helicosporidium* cyst suspension was poured into plugs until agarose polymerization. The plugs were then transferred into 10 mM TrisHCl containing 50 mM EDTA, 0.2% sodium deoxycholate, 1% lauryl succinate, and 1 mg/ml proteinase K and incubated at 37°C for 24h. After being washed four times in 50 mM EDTA at 37°C, the plugs were incorporated in a 1% agarose gel (in 0.5X TBE buffer). Intact chromosome electrophoresis was performed using a CHEF-DR II system (Biorad). The gel was run in 0.5X TBE buffer, at 6 V/cm for 24h, with a switching time ranging from 60 to 120 sec and stained in ethidium bromide.

DNA Extraction and PCR Amplification

Cellular DNA was extracted as previously described (Chapters 2 and 3), using the MasterPure Yeast DNA purification kit (Epicentre). The *Helicosporidium* sp. elongation factor gene *tufA* was amplified using the degenerate primers TufAf and TufAr (Appendix A). The resulting amplification product was gel-extracted and sequenced. Gene-specific primers (GSPs) were designed from the *Helicosporidium* sp. *tufA* sequence and used in

combination with primers designed from genes predicted to be located on a locus close to *tufA* within the chloroplast genome. The use of the fMET and rpl2R primers (Appendix A) allowed for the amplification and subsequent sequencing of the 5' and 3' flanking regions, respectively..

RNA Extraction and RT-PCR

Helicosporidium sp. cells were frozen under liquid nitrogen and ground into a fine powder. Total RNA was isolated using TriReagent, according to the manufacturer's protocol. To prevent any DNA contamination, *Helicosporidium* RNA was treated with RNase free DNase before being resuspended in formamide and stored at –70 °C. Prior to storage, an aliquot of the RNA suspension was used to spectrophotometrically estimate the final concentration. Upon utilization, stored RNA was reprecipitated in 4 volumes of 100% ethanol and 0.2M sodium acetate (pH=5.2) and suspended in distilled water. First-strand cDNA synthesis was performed using 1 µg of total RNA, the *tufA* gene specific primer LD PCR (see Appendix A for sequence), and the Thermoscript RT-PCR system from Life Technologies, following the manufacturer's directions. The LD PCR primer was then combined with a *rps12* and a *rps7* gene-specific primers in two separate reactions that were performed under the same conditions: 30 cycles of 94 °C for 30 sec., 50 °C for 30 sec, and 72 °C for 3 min.

Results

CHEF Gel Electrophoresis

The gel allowed for visualization of *Helicosporidium* sp. chromosomes (Fig. 4-1), suggesting that the cyst wall was disrupted by the treatment with DMSO and proteinase K. However, no bands corresponding to the mitochondrial or the plastid genomes were present (Fig. 4-1). Various modifications of the electrophoretic parameters were

performed, but they never resulted in any changes in the karyotype band pattern (data not shown). These results indicate that the circular chloroplast and mitochondrial DNA did not enter the gel, but remained in the well. Limited or no mobility for circular DNA molecules in CHEF gels has been reported previously (Higashiyama and Yamada, 1991; Maleszka, 1993) and have prevented from visualizing and estimating the size of the *Helicosporidium* sp. plastid genome. However, the CHEF electrophoresis provides information concerning the *Helicosporidium* sp. nuclear genome. This genome appears to be composed of 9 chromosomes, ranging from 700 kb to 2000 kb (Fig. 4-1). Summing up the sizes of individual chromosomal DNAs gave a 10.5 Mb estimate for the *Helicosporidium* sp. nuclear genome size. This estimate is much smaller than the genome size of its photosynthetic relative *Chlorella vulgaris* (estimated at 38.8 Mb; Higashiyama and Yamada, 1991).

Analysis of the Plastid Genome Sequence

Although the plastid DNA (ptDNA) was not observed on the CHEF gel, portions of this genome were readily PCR-amplified from *Helicosporidium* sp. total genomic DNA. A similar technique, based on the PCR amplification of overlapping sequences, was recently used to sequence the entire *Eimeria tenella* apicoplast genome (Cai et al., 2003). A 3348 bp fragment was amplified and sequenced from *Helicosporidium* sp. (GenBank accession number AY498714). Sequence comparison analyses demonstrated that the fragment contains four open reading frames (ORFs), corresponding to the elongation factor *tufA* and the ribosomal proteins *rps12*, *rps7*, and *rpl2*. In addition, the 5' end of the sequenced ptDNA fragment includes a portion of the proline tRNA (tRNA-P) gene. All five *Helicosporidium* sp. plastid genes are similar to homologous genes sequenced from

both *Prototheca wickerhamii* and *Chlorella vulgaris* chloroplast genomes. Furthermore, phylogenies reconstructed from a *tufA* alignment identified *Helicosporidium* sp. as a sister taxon to *Prototheca wickerhamii* (data not shown).

The overall organization of the sequenced *Helicosporidium* sp. ptDNA fragment is presented in Fig. 4-2. The *tufA*, *rps7* and *rps12* genes are known as the *str-* (streptomycin) cluster. This cluster is conserved across archeabacteria and eubacteria, including chloroplasts as intracellular descendants of the latter (Stoebe and Kowallik, 1999). Not surprisingly, the *str-* cluster is also conserved in *Helicosporidium* sp. plastid genome (Fig. 4-2). The *Helicosporidium* sp. ptDNA has an organization that is very similar to that *Prototheca wickerhamii*, especially in regard to the location of the *rpl2* gene. In both *Helicosporidium* sp. and *P. wickerhamii* ptDNA, this gene is located close to the 3' end of the *str-* cluster. This common organization differs from that of *Chlorella vulgaris* and other photosynthetic green algae (such as the ancestral *Nephroselmis olivacea*; Turmel et al., 1999), suggesting that the common ancestor of *Helicosporidium* sp. and *Prototheca wickerhamii* possessed a rearranged chloroplast genome. Rearrangements included the fusion of the *rlp2* cluster and *str-* cluster and may have been associated with the loss of photosynthesis.

Despite these similarities, the *Helicosporidium* sp. ptDNA fragment is also remarkably different from that of *Prototheca wickerhamii* (Fig. 4-2). First, two genes, corresponding to the ribosomal proteins *rpl19* and *rps23*, have not been found in *Helicosporidium* sp. As noted by Stoebe and Kowallik (1999), modifications in chloroplast genomes occur mainly in form of gene losses. Therefore, even if only a portion of the ptDNA has been sequenced, a likely hypothesis is that both *rpl19* and

rps23 have been lost from the *Helicosporidium* sp. plastid genome. Interestingly, a *rpl19* homologue has been identified in the Expressed Sequence Tag (EST) analysis of the *Helicosporidium* sp. nuclear genome (see Chapter 5). The consensus sequence obtained from two clones exhibited a 5' leader sequence that was found to be consistent with plastid targeting, suggesting that the *Helicosporidium* sp. *rpl19* gene may have been transferred from the plastid genome to the nuclear genome. In addition to the deletion of the *rpl19* and *rps23* genes, the orientation of the *str-* cluster in relation to the tRNA-P gene is different in *Helicosporidium* sp.: the tRNA-P gene is located on the same strand as the *str-* cluster and is transcribed in the same direction (Fig. 2). In contrast, the *Prototheca* tRNA-P orientation is similar to photosynthetic relatives such as *Chlorella vulgaris* and *Nephrolsemis olivacea*, suggesting that it represents an ancestral type among green algae. Overall, the *Helicosporidium* ptDNA fragment (Fig. 2) is characterized by a unique, derived organization, which may be the consequence of a genome rearrangement associated with gene losses and genome reduction.

RT-PCR Reactions

As presented in Fig. 4-3, the *str-* cluster was successfully amplified from *Helicosporidium* sp. cDNA, demonstrating that the ptDNA genes are expressed. Additionally, the RT-PCR products showed that the *str-* cluster genes are transcribed on the same mRNA molecule in an operon-like manner reminiscent of the chloroplast bacterial origin (Stoebe and Kowalllik, 1999). Importantly, the fact that plastid genes are expressed suggests that the *Helicosporidium* sp. plastid genome, despite being reorganized, has remained functional.

Discussion

Previous phylogenetic analyses (Chapters 2 and 3) have demonstrated that the Helicosporidia are close relatives of the non-photosynthetic algae *Prototheca* spp. (Chlorophyta; Trebouxiophyceae). In accordance with these analyses, *Helicosporidium* spp. are believed to possess a *Prototheca*-like plastid and a plastid genome (Chapter 3). Although the *Helicosporidium* sp. plastid has yet to be observed in microscopic examination, the combined PCR and RT-PCR amplifications presented in this study showed that *Helicosporidium* sp., as *P. wickerhamii*, has retained plastid genes, including the conserved *str-* cluster, that are expressed in helicosporidial cells. The presence of a transcribed ptDNA in *P. wickerhamii* has been demonstrated by Northern Blot analysis (Knauf and Hachtel, 2002). To date, the function of these vestigial organelles remains unclear.

A fragment of the *Helicosporidium* sp. ptDNA was sequenced and its architecture was compared to that of similar chloroplast genome fragments previously sequenced from both non-photosynthetic and photosynthetic relatives. These comparative genomic analyses revealed that the *Helicosporidium* sp. ptDNA is most similar to that of *Prototheca wickerhamii*, confirming that these two organisms arose from a common, recent ancestor (Chapters 2 and 3). However, a number of dissimilarities were also identified, suggesting that the Helicosporidia possess a unique, more derived plastid genome that has experienced additional gene losses and reorganization events. These observations indicate that the *Helicosporidium* sp. plastid genome may be more reduced than the 54 kb *Prototheca wickerhamii* ptDNA.

Concordant with the hypothesis that the helicosporidial ptDNA has been reduced in size is the fact that the nuclear genome appeared reduced as well. The *Helicosporidium* sp. nuclear genome has been estimated at 10.5 Mb (Fig 4-1), three times smaller than the genome of one of *Helicosporidium* sp. closest relatives, *Chlorella vulgaris* (38.8 Mb; Higashiyama and Yamada, 1991). Genome reduction is a common pattern observed for both pathogenic prokaryotes (Moran, 2002) and eukaryotes (Vivares et al., 2002), and it is always associated with the evolution toward pathogenicity and an obligate, host-dependent, minimalist lifestyle. Interestingly, biological observations that include the existence of a very specific infectious cyst stage (Boucias et al., 2001) and the ability to replicate intracellularly within insect hemocytes (Blaeske and Boucias, in press) have shown that the Helicosporidia possess characteristics that have not been reported for *Prototheca* spp. and that suggest that *Helicosporidium* spp. are more derived toward an obligate pathogenic lifestyle. Such observations concur with the hypothesis that the *Helicosporidium* sp. plastid genome may be smaller than that of *Prototheca wickerhamii*.

The generation of the complete sequence of the *Helicosporidium* sp. plastid genome will provide information on the extent of the genome reduction and rearrangement event(s). Potentially, the *Helicosporidium* sp. plastid genome is highly reduced, and may be more similar, in terms of size, gene content, and function, to the 35 kb apicoplast genome (Wilson, 2002) than to the 54kb *Prototheca wickerhamii* ptDNA. As noted by Williams and Keeling (2003), the Helicosporidia represent a remarkable opportunity to compare the evolution of non-photosynthetic plastids in two unrelated groups of intracellular pathogens. They may also prove to be a better model to study the transition from a free-living, autotrophic stage to a parasitic, heterotrophic stage and the

impact of this transition on both nuclear and plastid genomes (gene losses and transfers), because the phylogenetic affinity of *Helicosporidium* spp. and its relationships to both non-photosynthetic and photosynthetic relatives have been well established (Chapters 2 and 3), in contrast to the situation for Apicomplexa.

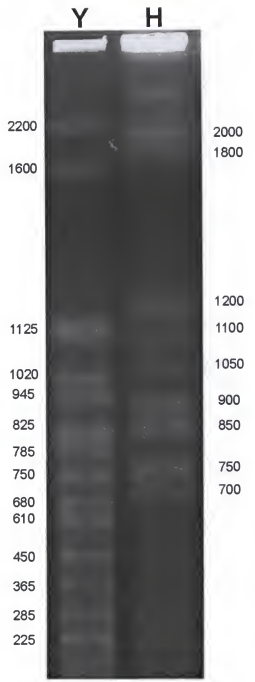


Figure 4-1: Karyotype analysis of the *Helicosporidium* sp. genome (H). The genome of the yeast *Saccharomyces cerevisae* (Y) was used as a reference to estimate the chromosome sizes (in kilobases). The absence of bands smaller than 700 kb suggests that the *Helicosporidium* sp. mitochondrial and plastid DNAs did not enter the gel, but remained in the well.

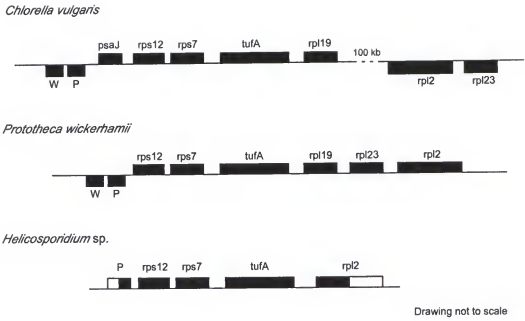


Figure 4-2: Comparison of the *Helicosporidium* sp. plastid genome fragment with that of non-photosynthetic (*Prototheca wickerhamii*) and photosynthetic (*Chlorella vulgaris*) close relatives. The sequenced regions are in black. The direction of transcription is from left to right for genes depicted above the lines and from right to left for those shown below the line.

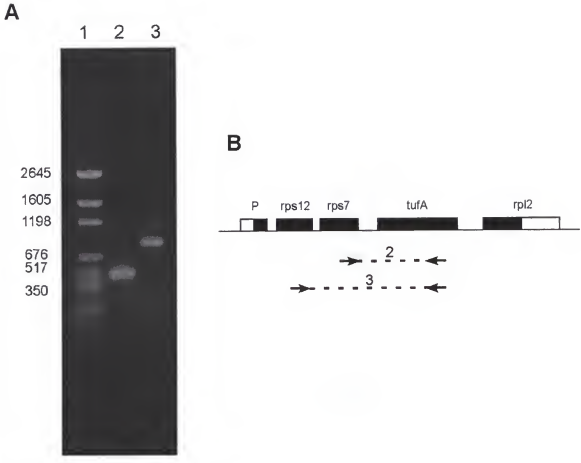


Figure 4-3: RT-PCR amplification of the *Helicosporidium* sp. *str*- cluster. (A) RT-PCR products run on a 1% agarose gel. The product in lane 2 was obtained using a combination of gene specific primers corresponding to the *rps7* (forward) and *tufA* (reverse) genes. The product in lane 3 was obtained with *rps12* (forward) and *tufA* (reverse) gene specific primers. DNA markers (pGEM) are shown in lane 1. (B) Schematic illustration of RT-PCR reactions.

CHAPTER 5
EXPRESSED SEQUENCE TAG ANALYSIS OF *HELICOSPORIDIUM* SP.

Introduction

The Helicosporidia are obscure pathogenic protists that have been reported in a wide range of invertebrate hosts (Keilin, 1921; Weiser, 1970; Kellen and Lindegren, 1973; Fukuda et al., 1976; Sayre and Clarke, 1978; Hembree, 1979; Purrini, 1984; Pekkarinen, 1993; Seif and Rifaat, 2001). Only one species of Helicosporidia has been described: *Helicosporidium parasiticum* Keilin 1921. To date, it remains unclear whether the group contains more than one species (see Appendix B) and whether these organisms are important insect pathogens and can be used as biocontrol agents against pest insects (Hembree, 1981; Seif and Rifaat, 2001).

Following the recent isolation of a new *Helicosporidium* sp. in Florida (Boucias et al., 2001), morphological and molecular data have been compiled on these little-known pathogens. Significantly, these data have demonstrated that the Helicosporidia are non-photosynthetic green algae, and they are related to *Prototheca*, another non-photosynthetic, parasitic algal genus (Boucias et al., 2001; Chapters 2 and 3). Several independent phylogenetic analyses showed that *Helicosporidium* sp. clusters within the class Trebouxiophyceae in a monophyletic clade that contains *Prototheca* spp. and *Auxenochlorella protothecoides*, suggesting that these organisms arose from a common ancestor (Chapters 2 and 3; also Ueno et al., 2003).

The reclassification of the Helicosporidia as green algae has ended an era of uncertainty in which *Helicosporidium* spp. were successively proposed to be Protozoa

(Kudo, 1931; Lindegren and Hoffman, 1976) or Fungi (Weiser, 1970) but were largely considered *incertae sedis* (Tanada and Kaya, 1993; Undeen and Vavra, 1997). Today, the Helicosporidia represent the only known entomopathogenic algae, but they remain very poorly characterized, especially at a molecular level. In an effort to better characterize the biology of the Helicosporidia, a large-scale sequencing project has been initiated by generating Expressed Sequence Tags (ESTs) from a *Helicosporidium* sp. cDNA library. EST sequencing has been recognized as a rapid, powerful, and cost effective method for genome analysis of eukaryotes. A large number of ESTs have been accumulated for a wide variety of organisms (see http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary .html for publicly available EST collections), including the chlorophytes *Chlamydomonas reinhardtii* and *Schefferlia dubia* (Asamizu et al., 1999; Becker et al., 2001; Shrager et al., 2003). However, no such large-scale sequencing effort ever has been reported for a green alga belonging to the class Trebouxiophyceae or for a non-photosynthetic green alga. The *Helicosporidium* sp. EST project described in this chapter consists of the accumulation of 1360 sequences, which increases significantly the very limited sequence information currently available for the Helicosporidia and provides insights into the biology of these unique organisms.

Materials and Methods

RNA Extraction

The *Helicosporidium* sp. isolated from the black fly *Simulium jonesii* (Boucias et al., 2001) was maintained on artificial media (TC insect medium supplemented by Fetal Calf Serum) and incubated at 26 °C. Cells were collected by low-speed centrifugation, resuspended into 10 ml of TriReagent (Sigma) plus glass beads (0.45 mm), and broken using a Braun MSK homogenizer. Following cell breakage, total RNA was extracted

using the TriReagent manufacturer protocol. Total RNA concentration was estimated spectrophotometrically. An aliquot of this resuspension was used to isolate polyA mRNA, using the Oligotex mRNA purification kit (Qiagen). PolyA mRNA was stored at –70 °C until cDNA synthesis.

Library Preparation and DNA Sequencing

The cDNA library was prepared in the Uni-ZAP XR plasmid using the ZAP-cDNA synthesis kit (Stratagene). Following the manufacturer's protocol, the cDNAs were ligated directionally into the Uni-ZAP XR vector, and the ligation reaction products were packaged using the Gigapack III Gold packaging extract. The library was then titered and amplified, and mass excision was performed in order to convert the phage into the pBluescript phagemid. *E. coli* colonies obtained after mass excision were screened by PCR for the presence of an insert and randomly transferred to 96-well plates. Plates were processed for sequencing both at the University of Florida (UF ICBR) and the University of British Columbia (UBC). Expressed Sequence Tags (ESTs) were obtained by single-pass sequencing of the 5' end of the cDNA clones using the T3 primer.

Sequence Analysis

The UF sequencing reads were imported in the ICBR software package "Finch-Suite" (by Geospiza Inc.) in which various third-party algorithms are used to estimate the quality of the read (Phred), trim down the vector sequences (Crossmatch), and assemble contigs (Phrap). ESTs obtained from UF and UBC, corresponding to fifteen (15) 96-well plates, were pooled into a common database. The non-readable sequencing reactions and vector-only reads were excluded from this database. Automated sequence similarity searches were done for each remaining EST using the BlastX algorithm to identify putative gene homologues in the non-redundant protein sequence database of the NCBI

(Altschul et al., 1990). BlastX E-values were used as a measure of sequence similarity, and ESTs with E-values < 10^{-5} were assigned to functional classes based on the functional catalog of plant genes (Bevan et al., 1998). Selected ESTs were also compared directly with the sequenced *Arabidopsis thaliana* genome (http://www.arabidopsis.org) and the *Chlamydomonas reinhardtii* genome (http://www.biology.duke.edu/chlamy/) using BLAST-inspired search engines available at these servers.

Phylogenetic Analyses

Consensus sequences from selected *Helicosporidium* sp. contigs were computationally translated, and the derived amino acid sequences were aligned with representative eukaryotic homologues (downloaded from GenBank) using ClustalX (Thompson et al., 1997). Single-gene datasets were combined to produce one concatenated amino acid alignment, and phylogenetic relationships were reconstructed using the parsimony and distance (Neighbor-Joining) methods implemented in PAUP* (Swofford, 2000).

Results

Features of the Generated ESTs

A total of 1360 clones were generated by random sequencing of a cDNA library from *Helicosporidium* sp. Similarity searches showed that half of these sequences (51.1%) do not possess any significant homologues in the NBCI non-redundant database (i.e., the BlastX E-value was higher than 10^{-5}).

The other half corresponds to 665 sequences with significant similarity to known sequences (E-values lower than 10^{-5}). A set of 387 contigs was assembled from these sequences (Fig. 5-1) and further analyzed. The 387 contigs represent unigenes, i.e., sequences that do not overlap with each other and, therefore, likely correspond to 387

genes. Most unigenes were represented by one single EST (282 unigenes out of 387), but a significant number of genes have been sequenced several times (Fig. 5-1). Among them, the genes encoding for the two subunits of the ribosomal DNA have the highest number of copies (more than 10) in the EST database (Fig. 5-1). A high proportion of the 387 contigs were shown to have very significant similarity to known protein sequences, with an E-value lower than 10^{-20} (Fig. 5-2). These high similarity values allowed for the assignment of both a closely related species and a putative function for each unigene. Therefore, the unigenes were classified according to the taxonomic distribution of their closest homologues (Fig. 5-3) and according to their functional categories (Fig. 5-4). These categories have been determined following the functional catalog of plant genes established for the analysis of the *Arabidopsis thaliana* genome (Bevan et al., 1998). Not surprisingly, green plants and green algae genes accounted for most of the matches (73%; Fig. 5-3), and most of the ESTs with similarity to known proteins were associated with typical interphase cell functions of a plant cell: assimilation of nutrients and biosynthesis of proteins (Fig. 5-4). The 387 *Helicosporidium* sp. unigenes, as well as their putative function, are listed in Table 5-1.

Significantly, 25% of the contigs are similar to protein sequences for which the function remains unclear or unknown, thereby lowering even more the final number of truly identifiable genes: 287 genes were identified with confidence out of our 1360 sequences. This low number of identifiable unigenes may be due, in part, to the uniqueness of *Helicosporidium* sp.

Phylogenetic Analyses of Conserved Proteins

Two unigenes were shown to be homologous to α-tubulin (clones 12G01 and 14A09) and to glyceraldehyde 3-phosphate dehydrogenase (GAPDH, clone 5F07). The contigs corresponded to the α-tubulin entire Open Reading Frame (ORF; 1350 bp), and a large fragment of the GAPDH ORF (606 bp). These two genes were selected for phylogenetic analyses because they encode for very conserved proteins and because a wide variety of homologous sequences are available in public databases. The two amino acid sequences were aligned with selected homologues. The alignments were combined and associated with the actin and β-tubulin amino acid sequence alignment (deduced from sequences obtained previously, see Chapter 2) to produce a concatenated, 1235 character alignment. The phylogenetic tree inferred from this data set is presented in Fig. 5-5. This tree includes several well-defined monophyletic eukaryote clades (Animals, Fungi, Green Plants, Green Algae, and Alveolates) and presents evolutionary relationships that correspond to the current consensus on eukaryotic phylogeny. Animals and Fungi are sister taxa. Alveolates are more closely related to the monophyletic clade formed by the green plants and algae (Viriplantae) than are the Opisthokonts (Animals and Fungi, see Chapter 1 for a review of eukaryotic current taxonomy). Importantly, the use of a large and informative concatenated alignment led to the fact that most of the nodes in the tree (including the deepest ones) are strongly supported by resampling tests (bootstrap). The tree depicts *Helicosporidium* sp. as a green alga, sister taxon to *Chlamydomonas reinhardtii*, with great confidence and confirms the results previously obtained throughout this study (Chapters 2, 3, and 4).

Identification of a Gene Possibly Acquired by Lateral Gene Transfer

Among the ESTs, two clones (2B11 and 6E01) were shown to exhibit significant similarities to bacterial proteases. The consensus contig sequence, inferred from an alignment of the two ESTs, is 678 bp long. PCR amplification and sequencing of a fragment of this consensus sequence has been performed (data not shown), confirming the helicosporidial origin of the protease gene. The deduced amino acid sequence of the *Helicosporidium* sp. protease was aligned with the closest homologues (according to BlastX analysis). Significantly, one of the closest relatives of the helicosporidial protease corresponds to an alkaline serine protease previously sequenced from the bacterial pathogen *Vibrio cholerae* (GenBank accession number NP_229814). The alignment of the two protein sequences is presented in Fig. 5-6. Similar alkaline proteases have also been cloned from other bacteria, including non-pathogenic species. Additionally, the *Helicosporidium* protease exhibits significant similarity to extracellular, cuticle-degrading proteases reported from various invertebrate pathogenic fungi, such as *Arthrobotrys oligospora* (PII protease; Ahman et al., 1996) and *Metarhizium anisopliae* (Pr1 protease; St Leger et al., 1992). These proteases are traditionally regarded as possible virulence factors. Therefore, the *Helicosporidium* protease also may be involved during the pathogenicity process.

Importantly, no homologous genes have been reported from algae or plants. Similarity searches within a plant (*Aradidopsis thaliana*) and a green alga (*Chlamydomonas reinhardtii*) genome did not reveal any clear plant-like homologues. In addition, the primers used to amplify the protease gene fragment from the *Helicosporidium* sp. genomic DNA failed to amplify a similar fragment from a

Prototheca zopfii genomic DNA preparation (data not shown). The protease gene exhibits a distinct phylogenetic signal, which is clearly different from that of the vast majority of the ESTs, suggesting that this gene might not have a plant/algal origin, but might have been acquired by *Helicosporidium* sp. via lateral gene transfer.

Discussion

A total of 1360 sequences have been produced from *Helicosporidium* sp. cDNA. From these, only 287 genes were identified with confidence. The fact that a large proportion of the *Helicosporidium* sp. ESTs could not be identified indicates that the Helicosporidia may harbor a large number of unique proteins. However, similar sets of data were previously obtained for two other algal EST projects involving the chlorophyte *Chlamydomonas reinhardtii* and the prasinophyte *Scherffelia dubia* (Asamizu et al., 1999, 2000; Becker et al., 2001). Both authors were surprised by the unexpectedly high number of unidentifiable sequences produced from two organisms that are known to be close relatives to land plants, for which extensive, and sometimes complete, genome sequence data are available. The number of unidentifiable sequences may reflect, in part, the uniqueness of these green algae, including *Helicosporidium* sp. However, Becker et al. (2001) also proposed that the lack of similarity may be explained by the fact that the genetic and phylogenetic heterogeneity within the Chlorophyta, as well as between chlorophytes and spermatophytes, may be much larger than previously expected. The complete sequencing of the *C. reinhardtii* nuclear genome will likely provide more information about the genetic and phylogenetic relationships between green plants and green algae. It also may help in identifying more *Helicosporidium* sp. genes, thereby strengthening this EST analysis. A complete molecular map of the *C. reinhardtii* genome

recently has been published (Kathir et al., 2003) and will be followed by a first-draft version of the complete genome sequence (http://www.biology.duke.edu/chlamy/).

Although the number of *Helicosporidium* sp. genes associated with known proteins was surprisingly low (387 unigenes), such sequence information provides insights into the biology of the poorly characterized Helicosporidia. Importantly, the overall phylogenetic signal of the ESTs (Fig. 5-3) demonstrates that *Helicosporidium* sp. has retained a plant-like cell metabolism. The identification of ca. 20 genes similar to nuclear-encoded, plastid-targeted genes (Keeling, personal communication) also provides indirect evidence that *Helicosporidium* sp. has conserved a plant-like cell organization, which includes a chloroplast-like organelle. A large number of these 20 ESTs exhibit a 5' leader sequence that is consistent with chloroplast targeting (Waller et al., 1998). The presence of a modified, but functional, chloroplast in Helicosporidia cells was previously demonstrated by the amplification of a chloroplast-like gene cluster from *Helicosporidium* sp. DNA preparations (Chapter 3 and 4). Lastly, phylogenetic analyses inferred from selected ESTs depicted *Helicosporidium* sp. as a member of the Plant eukaryotic supergroup (Baldauf, 2003). In summary, the sequence information provided by the EST analysis is consistent with the fact that the Helicosporidia are non-photosynthetic green algae.

In addition to the majority of plant-like genes, the ESTs also identified "foreign-looking" genes, including a bacteria-like protease. The Helicosporidia have evolved from a photosynthetic ancestor. However, losses of photosynthetic ability have appeared independently several times within the Chlorophyta, and most of the characterized non-photosynthetic green algae are not pathogenic. Therefore, the loss of photosynthesis does

not explain the *Helicosporidium* transition from an autotrophic to a parasitic stage. The identification of a bacterial gene provides possible evidence of lateral gene transfer and may explain this transition. As noted by de Koning et al. (2000), lateral gene transfer is the process by which genetic information is passed from one genome to an unrelated genome, where it is stably integrated and maintained. Lateral gene transfer between prokaryotes is a frequent and well-known phenomenon, but there has been accumulating evidence that this process also occurs between prokaryotes and eukaryotes and may be of particular importance in the evolution of a parasitic lifestyle (de Koning et al., 2000). Notably, acquisition of virulence factors from bacteria has been suggested for the entomopathogenic fungus *Metarhizium anisopliae* (Screen and St. Leger, 2000). The green alga *Helicosporidium* sp. may have acquired genes, including the protease gene, from unrelated organisms, and this acquisition may have led to the development of parasitism. Possibly, such genes have not been acquired, or conserved, by closely related organisms such as *Prototheca* spp. The complete sequencing of the protease gene, as well as thorough phylogenetic analyses, are currently underway and may confirm the gene transfer hypothesis and provide insights about the nature of the donor organism.

The trebouxiophyte *Helicosporidium* sp. is one of the few green algae for which a relatively large-scale sequencing effort has been developed. Similar molecular data have yet to be produced for *Helicosporidium* sp. closest relatives, such as *Chlorella vulgaris*, *Prototheca wickerhamii*, and *Prototheca zopfii*. Despite the relative lack of organisms suitable for comparative analyses, the EST database generated in this study provides a basis to study the cellular biology and the evolutionary history of the Helicosporidia.

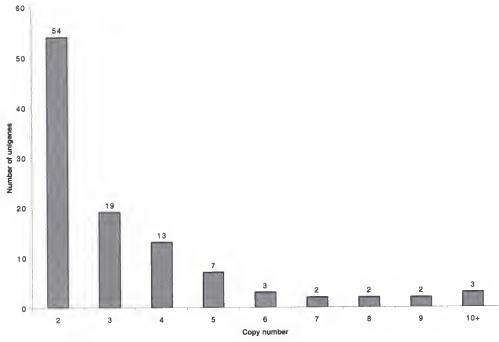


Figure 5-1: EST redundancy in contig assembly. While most of the unigenes are represented only once in the database (282 out of 387), some sequences are present twice or more. In this case, a consensus sequence (contig) has been computed.

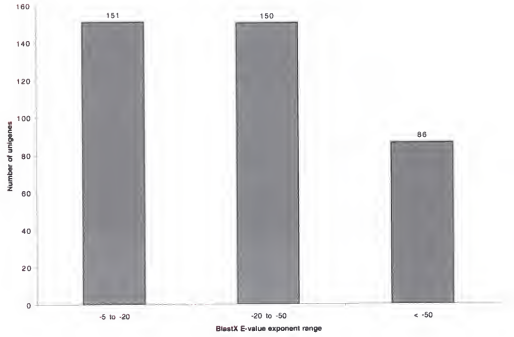
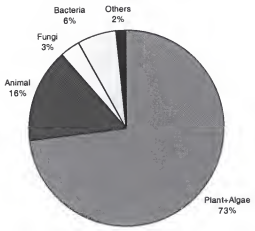


Figure 5-2: Sequence similarities between *Helicosporidium* sp. ESTs and the best match after BlastX analysis. The frequency of the resulting E-value is shown. A majority of unigenes (236 out of 387) exhibited significant similarity (with E-value lower than 10^{-20}), increasing the confidence that they have been correctly identified.

A



B

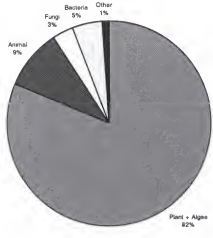


Figure 5-3: Taxonomic distribution of the closest homologues for the *Helicosporidium* sp. unigenes. (A) The 387 contigs with significant similarity to known proteins were classified according to the species the best BlastX match was sequenced from. Green plants and green algae accounted for most hits. (B) This distribution is clearer when only the 86 most similar contigs (E-value lower than 10^{-20}, see Fig. 5-2) are considered.

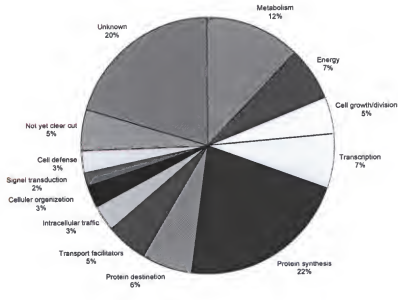


Figure 5-4: Functional classification of *Helicosporidium* sp. ESTs. The 387 unigenes were classified according to their putative function (determined by similarity searches via BlastX analyses)

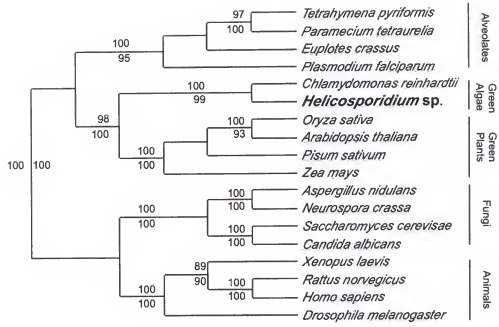


Figure 5-5: Phylogenetic (Neighbor-Joining) tree inferred from a concatenated alignment (1235 characters) containing four protein sequences corresponding to the actin, β-tubulin, α-tubulin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) genes. Numbers around the nodes correspond to distance (top) and parsimony (bottom) bootstrap values (100 replicates). The tree depicts *Helicosporidium* sp. as a green alga, with strong bootstrap support.

```
Helicosporidium sp.  ------------------------------------------------------------
Vibrio cholerae      MFKKFLSLCIVSTFSVAATSALAQPNQLVGKSSPQQLAPLMKAASGKGIKNQYIVVLKQP

Helicosporidium sp.  ----------MSDWSWPLINGTKDVHEPLRAYRVTGGLP-----LDARENKAQRVG----
Vibrio cholerae      TTIMSNDLQAFQQFTQRSVNALANKHALEIKNVFDSALSGFSAELTAEQLQALRADPNVD
                           :.:::   :*.  : *       . ..*.     * *.: :* *..

Helicosporidium sp.  ---------------------EELWSLDRIDQRSLPLDGYFNYGGASSAATGEGVVIY
Vibrio cholerae      YIEQNQIITVNPIISASANAAQDNVTWGIDRIDQRDLPLNRSYNYN-----YDGSGVTAY
                                         :  *.:******.***:  :**.       *.**. *

Helicosporidium sp.  VVDSGININHQEPQPFGGGPSRASYGYDFVDEDAEAADCDGHGTHVAASAAGLGVGVAKA
Vibrio cholerae      VIDTGIAFNHPEFG------GRAKSGYDFIDNDNDASDCQGHGTHVAGTIGGAQYGVAKN
                     *:*:** :** **       .**. ****:*:* :*:**:*******.: .*    ****

Helicosporidium sp.  ARVVAVRILDCSGSGSVTTTVAALDWVAAHAVKPAVVTLSLG----------------
Vibrio cholerae      VNLVGVRVLGCDGSGSTEAIARGIDWVAQNASGPSVANLSLGGGISQAMDQAVARLVQRG
                     ..:*.**:*.*.****. : . .:**** :*  *:*..****

Helicosporidium sp.  ----ISVGSWSKILAELAASRPHRGITGIPXCPWAIGANRRPWTA---------------
Vibrio cholerae      VTAVIAAGNDNKDACQVSPAREPSGITVGSTTNNDGRSNFSNWGNCVQIFAPGSDVTSAS
                         *:.*. .*  .:::.:*   ***  .        :*   *

Helicosporidium sp.  ------------------------------------------------------------
Vibrio cholerae      HKGGTTTMSGTSMASPHVAGVAALYLQENKNLSPNQIKTLLSDRSTKGKVSDTQGTPNKL

Helicosporidium sp.  ------------------------------------------------------------
Vibrio cholerae      LYSLTDNNTTPNPEPNPQPEPQPQPDSQLTNGKVVTGISGKQGELKKFYIDVPAGRRLSI

Helicosporidium sp.  ------------------------------------------------------------
Vibrio cholerae      ETNGGTGNLDLYVRLGIEPEPFAWDCASYRNGNNEVCTFPNTREGRHFITLYGTTEFNNV

Helicosporidium sp.  ------
Vibrio cholerae      SLVARY
```

Figure 5-6: Amino acid sequence alignment of the *Helicosporidium* sp. protease fragment with the homologous alkaline serine protease cloned from the pathogenic bacteria *Vibrio cholerae* (GenBank accession number NP_229814)

Table 5-1: List of the *Helicosporidium* sp. ESTs displaying significant amino acid similarity to the non-redundant GenBank protein database. The ESTs are classified according to broad cellular function.

Clone Ids	Putative function
Metabolism	
9H06	3 isopropylmalate dehydratase
3B04, 14C05	4 hydroxyphenylpyruvate
13E01	8 amino 7 oxononanoate synthase
7H02, 3B12	ACP stearoyl desaturase
12F06	acyl carrier protein (plastid)
11H11	acyl carrier protein (mitochondria)
5H12	adenosylhomocysteinase
4H05	adenylylsulfate kinase
2B11, 6E01	alkaline serine protease
13E10	beta-1,4-endoglucanase
4E04	beta mannase
3C04	proline dehydrogenase
2B02	oxysterol binding protein-like
4G08	cysteine proteinase
1A03	cysteine synthase
15C08	dihydroneopterin aldolase
4H10	putative 3-phosphoserine aminotransferase
1H03	2-isopropyl malate synthase
6B11	galactosidase beta1
3A12	glutathione-dependent formaldehyde dehydrogenase
9G07	oligoribonuclease
10C01	riboflavin kinase
3F09	glutamate-1-semialdehyde 2, 1-aminomutase
14C08	inosine-5'-monophosphate dehydrogenase
3D03	LYTB-like protein
3E08	NADP dependent steroid dehydrogenase
13F03, 10F09	nucleoside diphosphate kinase
10C04	cysteine proteinase precursor
14A08	UDP-Glucose 6 dehydrogenase
5B06	putative epimerase/dehydratase
8C12	hydrolase
1E11	molybdopterin synthase
5A10	UDP-N-acetylglucosamine pyrophosphorylase
9H07	riboflavin biosynthesis protein RibA
5B05	ribonuclease H related protein
7F07	S adenosylmethionine decarboxylase

Table 5-1. Continued

Clone Ids	Putative function
9B03	sterol-C5(6)-desaturase
12G06	sulfite synthesis pathway protein
6H03	intracellular protease/amidase protein (ThiJ family)
6E06	tyrosine carboxylase
15G06	UMP synthase
7D07	putative galactosyltransferase
12F04	Probable allantoinase
12A09	urate oxydase
Energy	
2B03	12-oxophytodienoate
13C02	aconitate hydratase
9F11	thioredoxin peroxydase
4D02	putative NADH dehydrogenase
10F08	putative aminotransferase (mitochondrial)
14D05	thioredoxin like
11H03	beta type carbonic anhydrase
15B10	cytochrome b5
9H04	cytochrome C1 precursor
4C08	putative lipoamide dehydrogenase
3B05	ferredoxin-thioredoxin reductase
13D01	fructose biphosphate aldolase
5F07, 15A03	glyceraldehyde 3-phosphate dehydrogenase
1D10	isocitrate dehydrogenase
3E10, 5G07, 2G04	malate dehydrogenase
5C03	NADP dependent malic enzyme
4E12	phosphoenolpyruvate carboxykinase
6B07	peroxiredoxin-like protein
3D07	phosphoglyceromutase
4H09	ubiquinol cytochrome c reductase
2A10	succinate dehydrogenase iron-sulfur subunit
14B10	succinate dehydrogenase subunit D
10F10, 14G10, 4D03, 8G08	Thioredoxin H
7F02	thromboxane A synthase (cytochrome P450 family)
8G07	Triosephosphate isomerase
5B12, 15A10	ubiquitin binding protein
Cell Growth/Division	
10A03, 10G04	DNA helicase-like

Table 5-1. Continued

Clone Ids	Putative function
11G09	flap endonuclease 1
4A06	Gbp1p telomere-associated protein
1D07, 6B08	guanine nucleide-binding protein
14B06	putative cell division protein FtsH protease-like
12H09	Centromere/microtubule binding protein
3G12	MAR-binding protein
3C10	DNA polymerase
6F06, 5A12	prohibitin
10E05, 5E03	proliferating cell nuclear antigen
4H11	protein kinase cdc2
4H02	Centromere/microtubule binding protein
7A06	nucleolar protein-like
6F08, 2D04	putative snRNP protein
15D10	ribonucleotide reductase large subunit B
11G12	spindle assembly checkpoint component
9G08	spindle pole body protein
1G01	Wd splicing factor
Transcription	
8F09	putative transcription factor
10H11, 3A12	26S ribosomal RNA
11F06	RNA helicase GU2
8B01	DNA-directed RNA polymerase II
3H04	RNA polymerase II subunit
2B08	glycyl tRNA synthetase
13C05	heterogeneous nuclear ribonucleoprotein
7F09, 15C05	histone H2B-I
7D09	histone H2B-IV
10B03, 15F02, 15F03	putative transcriptional coactivator
4E09, 2F12	polyadenylate-binding protein
4B02	RNA polymerase III
1A02	transcription factor tfIIH
7D04	RNA binding protein
3D06	putative RNA binding protein
6E05	splicing factor RSZ21
10C08	DNA directed RNA polymerase II largest subunit
B11	transcription factor hap5a-like
1E04	small nuclear riboprotein SmD1
4F05	nuclear RNA activating complex, polypeptide 3

Table 5-1. Continued

Clone Ids	Putative function
13A05	U6 snRNA-associated Sm-like protein
7E11	putative transcription factor APFI
11G01, 2B01	ribosomal protein S15
Protein Synthesis	
2H06	40S ribosomal protein S10
14A04, 13D06, 2H05, 6C06	40S ribosomal protein S11
9D05, 8H09, 10D01, 3F08, 7H04	40S ribosomal protein S16
10G10, 13D02, 12D02, 5H05, 14B04	40S ribosomal protein S19
10B08	40S ribosomal protein S2
13E07, 7G06	40S ribosomal protein S20
13D09, 12B11	40S ribosomal protein S21
13A10, 9H10, 13D03, 15B06, 6H01, 1H09, 4A04	40S ribosomal protein S23
14G03	40S ribosomal protein S24
12C01	40S ribosomal protein S3
7G02	40S ribosomal protein S8
14H03	40S ribosomal protein S9
1C09	50S ribosomal protein L15
6D01, 07H12, H07	5S ribosomal protein
2C06, 2A05, 10A02	60S acidic ribosomal protein P0
10H09, 15E04, 12B10, 13H08, 3A08, 12B07, 11H01	60S acidic ribosomal protein P1
9F01, 6A09, 8E10	60S acidic ribosomal protein P2
5C12	60S ribosomal protein L18
5E10, 5F11, 4C02	60S ribosomal protein L35
4A12, 1H08, 3E09, 1C10	60S ribosomal protein L10
4A02, 5B01, 13B11	60S ribosomal protein L11
4C05, 12B05, 12F11, 15H03	60S ribosomal protein L13
10F04	60S ribosomal protein L144
7H11, 12G07	60S ribosomal protein L15
10H06, 15F06, 15G01	60S ribosomal protein L17
2E11	60S ribosomal protein L18A
7C03	60S ribosomal protein L2
B08, 6D10, 13G02	60S ribosomal protein L21
11E02, 11H07	60S ribosomal protein L22
14D08, 10A06	60S ribosomal protein L23
07E03, 9E01	60S ribosomal protein L24
9D06, 15D12, 9H08, 8D11, 8B05	60S ribosomal protein L27

Table 5-1. Continued

Clone Ids	Putative function
1A04, 8E01, 13A08, 12A07	60S ribosomal protein L27A
2B09, 5D02, 08C08	60S ribosomal protein L28
6H08, 11C07, 8G05	60S ribosomal protein L31
10B07, 14C02	60S ribosomal protein L34
7B02	60S ribosomal protein L36-2
10D12	60S ribosomal protein L37
15D03, 6D08, 3A04, 12E06, 8D12, 1C04	60S ribosomal protein L37a
10B05, 9B08	60S ribosomal protein L38
6B06, 7H08	60S ribosomal protein L39
4F06, 13E08, 9B06	60S ribosomal protein L5
8B04, 3B10, 1C03, 7C02	60S ribosomal protein L6
11G04, 8A08, 7F11, 15B12	60S ribosomal protein L7A
8F03	putative translational inhibitor protein
4C11	40S ribosomal protein S13
6E04	ear1 protein
5G10, 4A03, 1A08, 7G05, 14D01, 13A04, 9D04, 12F07, 9C03	elongation factor 1 alpha long form
10F07, 7G07	elongation factor 2
15B05	nucleolar protein
14B12, 10A08	eukaryotic translation initiation factor 5A1
6C02	translation initiation factor 4E
10A04	translation initiation factor 4A
2A07	similar to 40S ribosomal protein S25
13E05	ribosomal protein L7a
6B10, 7D08, 3F11, 14E08, 8A04	ribosomal protein S29
3A07, 15C11, 1A05, 9C07, 7B04	ribosomal protein S28
2E02, 4A07, 13F04	hydroxyproline-rich ribosomal protein L14
3E03, 6B01, 1D09	initiation factor 5A
8C04, 13G06	methionyl-tRNA synthetase
12A10, 9H05	protein translation factor
13D05, 11C02	ribosomal protein S15
14F06	ribosomal protein SA (laminarin receptor)
2C02	40S ribosomal protein S3aA
1D12	50S ribosomal protein L33
1D06, 2C03, 13G11	60S ribosomal protein L23a
4G12, 13F05, 15F09	similar to plastid ribosomal protein L19
10A11, 04B03, 11F03	60S ribosomal protein L19
11A09, 14F05, 12C11, 15H11	60S ribosomal protein L26
B01, 4D12, 10B06, 2B07	ribosomal protein L9

Table 5-1. Continued

Clone Ids	Putative function
14H04, 12G10, 6D06, 6F12, 6G07	ribosomal protein S19 (S24)
14C06, 12E02	ribosomal protein S6
10D02	60S ribosomal protein L35A
8A07	translation initiation factor eIF-2B-delta subunit
3E05	tryptophanyl tRNA synthetase
6D11	translation initiation factor 2B beta subunit
15D04, 14A07, 15E06	ribosomal protein L30
10B10, 12B12, 10D03, 15E07, 15E05	ribosomal protein L32
15H04	ribosomal protein L7
5D08	ribosomal protein L8
14G01	ribosomal protein S14
10F01, 4E06, 11A08	ribosomal protein S26
2G09, 6E07, 1F04, 7D03, 13C06, 5F12	ribosomal protein S27
3E07, 2F02	ubiquitin extension protein/ribosomal protein S27a
Protein Destination	
3E12	26S proteasome ATPase subunit
7C09, 10F12	26S proteasome regulatory particle subunit 12
9G10	26S proteasome regulatory particle subunit 6
7B07	carboxypepdidase type III
5E08	protease II
3F12	serine carboxypeptidase-related
1H04	ADP ribosylation factor
11D02	putative chaperonine
5C05, 11A04	10 kDa chaperonine
5E01	putative signal recognition protein
9C09	FK506 binding protein-like
2F04	chaperonine 21 precursor
5F01	deoxyhypusine synthase
9G01	ubiquitin-conjugating enzyme 1
4H03	peptidyl-prolyl cis-trans isomerase
6C10, 7B05	peptidylpropyl isomerase
13H10, 5E09	phosphomannomutase
4D11, 9A02, 1G11	polyubiquitin
15D08	aminopeptidase N metalloprotease
11H05, 10C03	prolyl 4-hydrolase alpha subunit
6A03, 1F03, 8D08, 14C12	protein disulfide isomerase
10B01	ubiquitin activating enzyme E1C

Table 5-1. Continued

Clone Ids	Putative function
14F01	T complex protein 1 epsilon subunit
7A08	ubiquitin conjugating enzyme
3H05	ubiquitin conjugating enzyme
4D10	ubiquitin conjugating enzyme
12F12	putative prolylcarboxypeptidase
Transport Facilitators	
12E11, 10E01	ADP-ATP carrier protein
14E11	amino acid permase AAP3
7E08	aminoacid permase AAP5
3G03	cis-Golgi SNARE protein
12G02	coatamer alpha subunit
2G06	copper chaperone homologue
10G05	epsilon subunit of mitochondrial F1-ATPase
14C11	glucose-6-phosphate/phosphate translocator
3D05, 15G11	ferredoxin
2A09	Pi transporter homologue
15A07	Plasma membrane ATPase
11D04	porin-like protein
1G12	ABC transporter subunit
11H10	ATP synthase delta chain
10H10, 12H10	coatmer beta subunit
9C01	H+ transporting ATP synthetase
1F10	probable transaminase
13G08	phosphate/phosphosenolpyruvate translocator
4B01	vacuolar ATP synthetase subunit F
2B10	vacuolar ATP synthetase subunit B
Intracellular Traffic	
13B08	cytochrome P450
12D05	synaptobrevin-like
1A07	GTP-binding protein yptV5
4F08	GTP-binding protein yptV1
4C10, 5F03	Ligatin
8A02	mitochondrial carrier like protein
9B02	mitochondrial 2 oxoglutarate/malate translocator
4B10	GTP-binding protein SAR1
13G10	GTP-binding protein
10D09	synaptobrevin-like

Table 5-1. Continued

Clone Ids	Putative function
7B03	signal recognition particle 54 kDa (SRP54)
8B08	signal recognition particle 19 kDa
12H07	mitochondrial uncoupling protein
Cellular Organization	
11C05	beta expansin
7C07	mitochondrial 23S rDNA
8H12	phosphatidylserine receptor
4E07	profilin
12B08	cell wall-bound apyrase
12E05	cytoskeleton associated protein
11B02	JUN kinase activator protein
11G07	ribophorin-I homologue
12H08, 7D06	sperulin 1b
14A09, 12G01	Tubulin alpha chain
Signal Transduction	
10A10	calmodulin binding structure
2F07, 15H06	calmodulin
13E11	casein kinase
3C01	calcium binding protein
14D04	MAP kinase phosphatase
6E03	protein kinase ck2 alpha subunit
8D03	protein kinase ck2 regulatory (beta) subunit
Cell Defense	
9F05	chymotrysin inhibitor 2
12B06, 13D04	glycine-rich protein 2
2C07	heat shock cognate protein
1E05	heat shock protein 70
4F09	heat shock protein 90
6F05, 3C12, 6G10, 2C12, 4C09, 4F11, 3D08, 3A03, 9C08, 12H12, 13B05, 14H10, 14H01, 10C09, 13A01, 10H08	heat shock protein 20
3D04	ClpB heat shock protein-like
15C04	similar to fungal resistance protein
07E01	putative glutathione peroxidase
1D11	metallothionein

Table 5-1. Continued

Clone Ids	Putative function
Not Yet Clear Cut	
15G08	anti-silencing function 1a protein
6G05	putative cap binding protein
9E10	cleft lip and palate associated transmembrane protein
3A11	rhodanese-like family protein
7C11	CsgA protein
12D06	glycine hydroxymethyltransferase
2A04	hyuC-like protein
15E12	leucine-rich repeat transmembrane protein kinase
9B04	expressed protein (rhs)
12B09	ovarian abundant message protein
6F03	carboxymethylenebutenolidase
9H09	putative esophageal gland cell secretory protein
5G05, 1E10, 6C05, 11A10, 15G04, 15B08, 10E11, 8D10	putative regulatory protein
6B04, 7G03	putative senescence-associated protein
11G11	putative transmembrane protein
4B06	selenium binding protein
15H05	senescence associated protein
7H03, 4D09	stress-induced protein sti1
12H01	testis expressed gene 261
4C06	MCT-1 protein-like
13H03	zygote specific protein
Unknown	
10F05	Hypothetical protein (EST anopheles)
7A03	Hypothetical protein (EST anopheles)
13C03	Hypothetical protein (EST anopheles)
8C02	putative protein
14C10, 13E06	hypothetical protein
9G11	B12D protein
8G12	hypothetical protein
6E09	expressed protein
14H11	expressed protein
10B04	expressed protein
10D06	expressed protein
1F09	expressed protein
14A11	expressed protein

Table 5-1. Continued

Clone Ids	Putative function
15F07	expressed protein
15B11	expressed protein
15B03	expressed protein
14E07	expressed protein
15E01	expressed protein
13C08	expressed protein
10D07	expressed protein
10G13	expressed protein
14D02	expressed protein
10H04	expressed protein
11G10	expressed protein
5E11	expressed protein
5E02	expressed protein
4F07	expressed protein
2G02, 7D02	hypothetical protein
11E01	hypothetical protein
1B09	expressed protein
6G01	expressed protein
15G10	hypothetical protein
7G09	expressed protein
12F08	hypothetical protein
07F03	hypothetical protein
12D04	hypothetical protein
10G02	hypothetical protein
11E08	hypothetical protein
9B11	acyl CoA binding protein, putative
5D11	hypothetical protein
14G04	hypothetical protein
10D04, 12A01, 11B07	hypothetical protein
4D07	hypothetical protein
7A11	hypothetical protein
1E07	hypothetical protein
15H01	ORF1 - putative transposase
10A01	hypothetical protein
14F04	hypothetical protein
7B06	hypothetical protein
8G06	hypothetical protein
15G12	putative protein
15B07	pollen specific protein

Table 5-1. Continued

Clone Ids	Putative function
1A06	hypothetical protein
1G04	hypothetical protein
4G04	hypothetical protein
7C08	hypothetical protein
13G08	hypothetical protein
6C01	hypothetical protein
12D08	hypothetical protein
10G11	hypothetical protein
10D09	hypothetical protein
HEL11E04	hypothetical protein
4G07	hypothetical protein
2A06	hypothetical protein
4G03	hypothetical protein
5E12	hypothetical protein
14G05	hypothetical protein
3A02	expressed protein
14E03	expressed protein
6D02	expressed protein
09D03, 14A05	expressed protein
7D01	expressed protein
3H08, 11F05, 8G09, 11F02, 8D04, 4G11	expressed protein
11E07	expressed protein
8B07	expressed protein
5E05	expressed protein
11D07	expressed protein
9E11	expressed protein
11C03, 5G01	expressed protein
Transposons	
7H01	putative polyprotein (retroelement)

CHAPTER 6
SUMMARY AND DISCUSSION

This study presents the first molecular sequence comparison analyses that include the genus *Helicosporidium*. Surprisingly, these analyses have recurrently identified the Helicosporidia as green algae (Chlorophyta). This taxonomic position never has been suggested by previous studies on *Helicosporidium* spp., which associated these organisms either with fungi or protozoa (see literature review in Chapter 1). Phylogenetic analyses, coupled with cellular biology evidence (presence of a chloroplast) and morphological evidence (the peculiar growth of *Helicosporidium* sp.; see Boucias et al., 2001), have demonstrated that the Helicosporidia are the first described entomopathogenic green algae. Furthermore, in contrast to most previous *Helicosporidium* taxonomic classification attempts, this study associated the Helicosporidia with other known protists: the non-photosynthetic green algae *Prototheca* spp. (Chlorophyta, Trebouxiophyceae).

Evolutionary History of the Helicosporidia

Both phylogenetic analyses (Chapters 2 and 3) and plastid genome comparisons (Chapter 4) presented in this study have shown that the genera *Helicosporidium* and *Prototheca* are very close relatives and have evolved from a common ancestor. The plastid *rrn16* phylogeny (Chapter 3) identified *Helicosporidium* spp. as a member of the *Prototheca* clade (Nedelcu, 2001), which is composed exclusively of non-photosynthetic, unicellular green algae *Prototheca* spp., except for the photosynthetic *Auxenochlorella protothecoides* (Nedelcu, 2001).

The *Helicosporidium-Prototheca* relationship that has been demonstrated throughout this study has since been confirmed by another independent analysis (Ueno et al., 2003). Although it is clear that *Auxenochlorella protothecoides*, *Prototheca* spp. and *Helicosporidium* spp. form a monophyletic clade (this study; Huss et al. 1999; Nedelcu, 2001; Ueno et al., 2003), the relationships within this clade have yet to be resolved. As noted by Ueno et al. (2003), very limited sequence information has been gathered for *Prototheca* spp., which has restricted the extent of previous phylogenetic analyses that included the *Prototheca* clade. Significantly, the genus *Prototheca* is always paraphyletic. In this study and in others, *P. wickerhamii* consistently is depicted as more closely related to the photosynthetic *A. protothecoides* than to *P. zopfii* (see Chapter 2; Nedelcu, 2001; Ueno et al., 2003). When included, *Helicosporidium* spp. are depicted as sister taxa to *P. zopfii* (Chapter 2 and 3; Ueno et al., 2003). SSU and LSU rDNA phylogenies also associated the other *Prototheca* spp. (*P. ulmea*, *P. stagnora*, and *P. moriformis*) with *P. zopfii* and *Helicosporidium* sp. (Ueno et al., 2003).

Because of the apparent paraphyletic nature of the genus *Prototheca*, no single most parsimonious *Helicosporidium* evolutionary scenario may be advanced, and the exact occurrence of the loss of photosynthesis remains unclear (Fig. 6-1). As noted by Huss et al. (1999), it would be more parsimonious if *Auxenochlorella protothecoides*, which is photosynthetic, were ancestral to all non-photosynthetic species. In all phylogenetic analyses performed to date, this is never the case, and two scenarios remain (Fig. 6-1). The first one involves one single loss of photosynthesis, experienced by the common ancestor to *A. protothecoides*, *Prototheca* spp., and *Helicosporidium* spp. This scenario implies the reappearance of autotrophy for *A. protothecoides*, but is consistent

with the fact that this species is auxotrophic and mesotrophic (Huss et al., 1999; also discussed by Nedelcu , 2001). The alternative scenario involves two independent losses of photosynthesis for both *Helicosporidium* sp. and *Prototheca wickerhamii* (Fig. 6-1).

The evolution of parasitism is likely to be specific to the Helicosporidia, as they are the only organisms in the *Prototheca* clade that are associated with invertebrates. Additionally, *Prototheca wickerhamii* and *Prototheca zopfii* are only mild pathogens, and the other *Prototheca* spp. are not known to be pathogenic or even, in the case of *P. stagnora*, associated with animals (Pore, 1985). As stated in Chapter 5, one likely hypothesis is that the *Helicosporidium* spp. ancestor has acquired genes that would enable it to become pathogenic to an invertebrate host. These genes must not have been acquired or conserved by *Prototheca* spp., leading to the separation of the two genera. However, this idea remains largely a hypothesis, and the exact number and nature of transferred genes, as well as the nature of the donor organism(s), have yet to be resolved.

The phylogenetic analyses presented in this study allow hypotheses about the evolution of the non-photosynthetic algae *Helicosporidium* spp. from a photosynthetic ancestor common to the *Prototheca* clade to be put forth and tested. The relationships within this clade may be resolved by producing additional sequence data, especially from poorly characterized organisms such *as Auxenochlorella protothecoides* and *Prototheca zopfii*. Although their evolution remains largely unresolved, it is clear that the Helicosporidia are non-photosynthetic green algae and unique invertebrate pathogens.

The Helicosporidia Reflect the Entomopathogenic Protist Diversity

As stated above, the Helicosporidia, now identified as non-photosynthetic green algae, represent a new type of entomopathogenic eukaryote. Insect pathogenic protists

have evolved independently within several major eukaryotic groups (Table 6-1) and now have been reported in at least six of the eight supergroups identified by Baldauf (2003). In some eukaryotic lineages, such as the fungi, entomopathogenic organisms have appeared independently several times. Most of these organisms, and especially their pathogenic strategies, remain very poorly known. However, the fact that numerous entomopathogenic eukaryotes have appeared within distinct eukaryote groups suggests that they may have evolved different pathogenic strategies. Entomopathogenic protists include intracellular and extracellular pathogens, illustrating the wide variety of strategies that are known to be used by these organisms. To date, these strategies are understudied and underexploited. Only a few entomopathogenic eukaryotes are being developed as effective biocontrol agents (i.e., *Metarhizium anisopliae* and *Beauveria bassiana*; see Butt et al., 2001), and their use is extremely restricted, especially when compared to other types of insect pathogens, such as viruses, bacteria, or nematodes.

The entomopathogenic eukaryotes (traditionally considered as Protozoa) are the least understood entomopathogens. The Helicosporidia, after being correctly identified as non-photosynthetic green algae nearly 100 years after their first discovery, exemplify both our limited knowledge on insect pathogenic eukaryotes and the potential these eukaryotes represent as novel biocontrol agents.

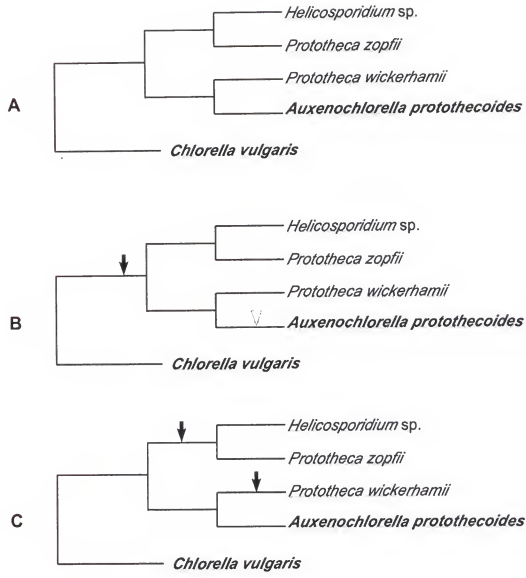


Figure 6-1: Evolutionary scenarios for *Helicosporidium* sp. (A) Consensus phylogenetic relationships within the *Prototheca* clade. The photosynthetic species are in bold. (B) One most parsimonious scenario involves one loss of photosynthesis (black arrow) and one reappearance of autotrophy (white arrow). (C) Another equally parsimonious scenario involves two independent losses of photosynthesis (black arrows).

Table 6-1: List and taxonomic affiliations of entomopathogenic eukaryotes.

Eukaryotic groups	**Subgroups**	**Genera**
Opistokhonts	Fungi: Chytrids	*Coelomomyces*
	Fungi: Microsporidia	*Nosema, Vairimorpha*
	Fungi: Zygomycetes	*Entomophthora*
	Fungi: Ascomycetes	*Metarhizium, Beauveria*
Amoebozoa	-	*Malamoeba, Malpighamoeba*
Plants	Chlorophyta	*Helicosporidium*
Alveolates	Apicomplexa	*Ascogregarina, Mattesia*
	Ciliates	*Lambornella*
Heterokhonts	Oomycetes	*Lagenidium*
Discicristates	Kinetoplasts	*Leptomonas*
Incertae sedis	-	*Nephridiophaga*

APPENDIX A
LIST OF PRIMERS USED IN THIS STUDY

Table A-1: List of primers used to PCR-amplify *Helicosporidium* spp. nuclear genes. Also indicated are the primer sequences and amplification conditions.

Genes & Primer Information	Tm	Est. fragment size	Comments
18S rDNA			
Forward:			
18S69F - CTGCGAATGGCTCATTAAATCAGT	55 °C	69F-1118R: 1000 bp	
18S363F – CGGAGAGGGAGCCTGAGAAA		363F-1577R: 1200 bp	
Reverse:		69F-1577R: 1500 bp	
18S1118R – GGTGGTGCCCTTCCGTCAA			
18S1577R – CAAAGGGCAGGGACGTAATCAA			Combination with 18S primers are possible
Gene-specific:			
HelicoSSU_F – ACACGAGGATCAATTGGAGGGC	55 °C	SSU_F-SSU_R: 400 bp	
HelicoSSU_R - CAATGAAATACGAATGCCCCCG			
28S rDNA			
Forward:			
D1/D2-NL4 - GGTCCGTGTTTCAAGACGG	55 °C	NL1-NL4: 680 bp	
Reverse:			
D1/D2-NL1 - GCATATCAATAAGCGGAGGAAAAG			
5.8S rDNA			
Forward:			
TW81 – GTTTCCGTAGGTGAACCTGC	55 °C	TW81-AB28: 950 bp	
Reverse:			
AB28 - ATATGCTTAAGTTCAGCGGGT			
Actin			
Forward:			
ED35 – CACGGYATYGTBACCAACTGGG		ED35-ED30: 800 bp	
ED33 – TTCGAGACHTTCAACGTSCC		ED33-ED30: 700 bp	
ED31 – GAAACTACCTTCAACTCCATCATG	50 °C	ED31-ED30: 300 bp	Also work on fungal DNA
Reverse:			
InvED31 – CTTGCGGATGTCCACGTCG		ED35-InvED31: 500 bp	
ED30 - CTAGAAGCATTTGCGGTGGAC		ED33-InvED31: 400 bp	
β-Tubulin			
Forward:			
TubF - TGGGCYAARGGYCACTACACYGA			Also work on fungal DNA
Reverse:	55 °C	TubF-TubR: 900 bp	
TubR - TCAGTGAACTCCATCTCRTCCAT			

Table A-2: List of primers used to PCR-amplify *Helicosporidium* spp. mitochondrial genes. Also indicated are the primer sequences and amplification conditions.

Genes and Primer Information	Tm	Est. fragment size	Comments
Cox3 Forward: CC66 – GTAGATCCAAGTCCATGG Reverse: CC67 - GCATGATGGGCCCAAGTT	50 °C	CC66-CC67: 400 bp	

Table A-3: List of primers used to PCR-amplify *Helicosporidium* spp. plastid genes. Also indicated are the primer sequences and amplification conditions.

Gene and Primer Information	Tm	Est. fragment size	Comments
16S rDNA Pair #1: ms-5' - GCGGCATGCTTAACACATGCAAGTCG ms-3' - GCTGACTGGCGATTACTATCGATTCC Pair #2: rrn16F - AGTRGCGRACGGGTGAGTAA rrn16R - GACARCCATGCACCACCTGT	50 °C 50 °C	ms-5'-3': 1200 bp rrn16F-R: 900 bp	ms primers from Nedelcu (2001) J. Mol Evol. rrn16 primers are not suitable for sequencing
tufA Forward: TufAf – AAYATGATTACAGGTGCTGC Reverse: TufAr - ACGTAAACTTGTGCTTCAAA	50 °C	TufAf-r: 700 bp	
Plastid genome fragment fMET – GGGTAGAGCAGTCTGGTAGC rpl2R - CCTTCACCACCACCATGCG	50 °C	3.5 kb	

APPENDIX B
A SECOND *HELICOSPORIDIUM* SP. ISOLATE

During my studies on the *Helicosporidium* sp. isolate found in a black fly larva, a second isolate has been identified. It has been isolated from the weevil *Cyrtobagous salviniae* (Coleoptera: Curculionidae). This insect is a biological control agent for the aquatic weed *Salvinia molesta* (Goolsby *et al.*, 2000). The two isolates will be referred to as weevil *Helicosporidium* and black fly *Helicosporidium*.

The weevil *Helicosporidium* was successfully amplified in *Helicoverpa zea* larvae as well as in artificial media. Following the protocols established for the black fly *Helicosporidium*, DNA extraction also has been performed. Most of the gene amplifications reported in this study have been duplicated using the weevil *Helicosporidium*, and sequences corresponding to the SSU rDNA, actin, β-tubulin, mitochondrial *cox3*, and plastid *rrn16* have been used in comparative analyses. Phylogenetic trees that include both *Helicosporidium* isolates are presented in Figs. B-1 through B-4. In these trees, the Helicosporidia are always depicted as a monophyletic group. However, the two *Helicosporidium* isolates exhibit some polymorphism in all sequenced genes, suggesting that they can be differentiated at a molecular level.

Based on morphological comparisons, Lindegren & Hoffman (1976) introduced the hypothesis that there may be more than one species of *Helicosporidium*. Here, it remains unclear whether the observed nucleotide differences are significant and sufficient to propose that the black fly and weevil *Helicosporidium* represent different strains or species. A thorough characterization of these two isolates is currently underway.

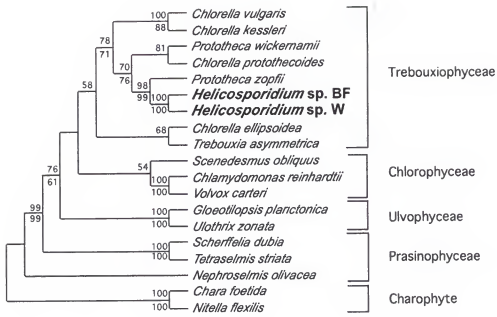


Figure B-1: Phylogenetic tree (Neighbor-Joining) inferred from a SSU rDNA alignment. The tree includes both *Helicosporidium* isolates, depicted as a monophyletic group sister taxa to *Prototheca zopfii*. The letters W and BF respectively refer to the weevil and the black fly *Helicosporidium*. Numbers around the nodes correspond to bootstrap values (100 replicates) obtained with distance (top) and parsimony (bottom) method. Only values greater than 50% are shown.

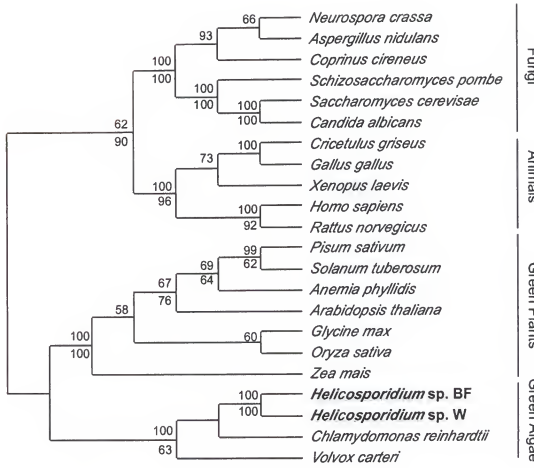


Figure B-2: Phylogenetic tree (Neighbor-Joining) inferred from a concatenated dataset that included both actin and β-tubulin nucleotide sequences. The two *Helicosporidium* isolates group within the green algae. The letters W and BF respectively refer to the weevil and the black fly *Helicosporidium*. Numbers around the nodes correspond to bootstrap values (100 replicates) obtained with distance (top) and parsimony (bottom) method. Only values greater than 50% are shown.

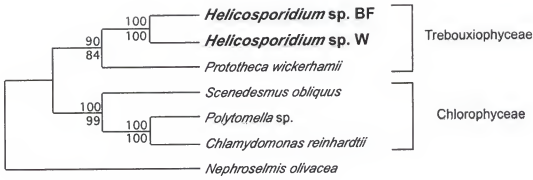


Figure B-3: Phylogenetic tree inferred from a *cox3* amino acid sequence alignment. The tree shows that *Helicosporidium* and *Prototheca* are closely related genera. The letters W and BF respectively refer to the weevil and the black fly *Helicosporidium*. Numbers around the nodes correspond to bootstrap values (100 replicates) obtained with distance (top) and parsimony (bottom) method. Only values greater than 50% are shown.

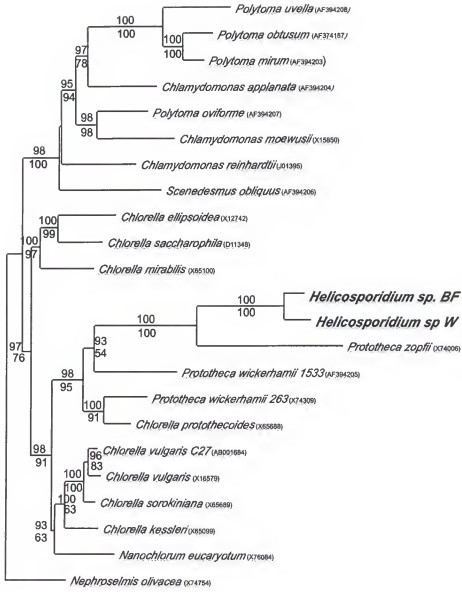


Figure B-4: Phylogram inferred from a plastid *rrn16* alignment. Once again, the two *Helicosporidium* isolates cluster together as a monophyletic group. This group is included into a strongly supported *Prototheca* clade (sensu Nedelcu, 2001) that clusters *Helicosporidium* spp., *Prototheca* spp. and *Chlorella protothecoides*. The letters W and BF respectively refer to the weevil and the black fly *Helicosporidium*. Numbers around the nodes correspond to bootstrap values (100 replicates) obtained with distance (top) and parsimony (bottom) method. Only values greater than 50% are shown.

APPENDIX C
ACCESSION NUMBERS FOR HELICOSPORIDIAL SEQUENCES

Table C-1: GenBank accession numbers affiliated with the *Helicosporidium* spp. nucleotide sequences obtained in this study.

	Black fly *Helicosporidium*	Weevil *Helicosporidium*
SSU rDNA (18S)	AF317893	-
LSU rDNA (28S)	AF317894	-
ITS1-5.8S-ITS2	AF317895	-
Actin	AF317896	-
Beta-tubulin	AF317897	-
Mitochondrial *cox3*	AY445515	AY445516
Plastid SSU rDNA (16S)	AF538864	AF538865

LIST OF REFERENCES

Ahman, J., Ek, B., Rask, L. & Tunlid A. (1996). Sequence analysis and regulation of a gene encoding a cuticle-degrading serine protease from the nematophagous fungus *Arthrobotrys oligospora*. *Microbiology* **142,** 1605-1616.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215,** 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acid Res* **25,** 3389-3402.

Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. & Tabata, S. (1999). A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant Expressed Sequence Tags. *DNA Res* **6,** 369-373.

Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K, Nakamura, Y. & Tabata, S. (2000). Generation of expressed sequence tags from low CO2 and high-CO2 adapted cells of *Chlamydomonas reinhardtii*. *DNA Res* **7,** 305-307.

Avery, S. W. & Undeen, A.H. (1987a). The isolation of Microsporidia and other pathogens from concentrated ditch water. *J Am Mosq Control Assoc* **3**, 54-58.

Avery, S. W. & Undeen, A. H. (1987b). Some characteristics of a new isolate of *Helicosporidium* and its effect upon mosquitoes. *J Invertebr Pathol* **49,** 246-251.

Baldauf, S. L. (2003). The deep roots of eukaryotes. *Science* **300,** 1703-1706.

Baldauf, S. L. & Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci USA* **90,** 11558-11562.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290,** 972-977.

Becker, B., Feja, N. & Melkonian, M. (2001). Analysis of Expressed Sequence Tags (ESTs) from the scaly green flagellate *Scherffelia dubia* Pascher emend. Melkonian et Preisig. *Protist* **152,** 139-147.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S.A., Patel, K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T.M., Terryn, N., Gielen, J., Villarroel, R. & Chalwatzis, N. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391,** 485–488.

Bhattacharya, D. & Medlin, L. (1998). Algal phylogeny and the origin of land plants. *Plant Physiol* **116,** 9-15.

Boucias D.G. & Pendland, J.C. (1998). Principles in insect pathology. Kluwer Academic Publishers, Boston.

Boucias, D. G., Becnel, J. J., White, S. E. & Bott, M. (2001). In vivo and in vitro development of the protist *Helicosporidium* sp. *J Eukaryot Microbiol* **48,** 460-470.

Butt, T.M., Jackson, C.W. & Magan, N. (2001). Fungi as biocontrol agents. Progress, problems and potential. CABI Publication.

Cai, X., Lorraine Fuller, A., McDouglas, L.R. & Zhu, G. (2003). Apicoplast genome of the coccidian *Eimeria tenella*. *Gene* **321,** 39-46.

Cavalier-Smith, T. (1993). Kingdom Protozoa and its 18 phyla. *Microbiol rev* **57,** 953-994

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol Rev Cambridge Phil Soc* **73,** 203-266.

Cavalier-Smith, T. & Chao, E.E.-Y. (2003). Phylogeny of Choanozoa, Apusozoa, and other protozoa and early eukaryote megaevolution. *J Mol Evol* **56,** 540-563.

Curran, J., Driver, F., Ballard, J. W. O. & Milner, R. J. (1994). Phylogeny of *Metarhizium*: analysis of ribosomal DNA sequence data. *Mycol Res* **98,** 547-552.

De Koning, A.P., Brinkman, F.S.L., Jones, S.J.M. & Keeling, P.J. (2000). Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol Biol Evol* **17,** 1769-1773.

Drouin, G., Moniz de Sa, M. & Zucker, M. (1995). The *Giarda lamblia* actin gene and the phylogeny of eukaryotes. *J Mol Evol* **41,** 841-849.

Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. (1994). Testing significance of incongruence. *Cladistics* **10,** 315-319.

Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D. & Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12,** 119-124.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* **27,** 401-410.

Fukuda, T., Lindegren, J. E. & Chapman, H. C. (1976). *Helicosporidium* sp. A new parasite of mosquitoes. *Mosquito News* **39,** 514-517.

Galan, F., Garcia-Martos, P. ,Palomo, M. J., Beltran, M., Gil, J. L. & Mira, J. (1997). Onychoprotothecosis due to *Prototheca wickerhamii*. *Mycopathologia* **137,** 75-77.

Gockel, G. & Hachtel, W. (2000). Complete gene map of the plastid genome of the non photosynthetic euglenoid flagellate *Astasia longa*. *Protist* **151,** 347-351.

Goolsby, J. A., Tipping, P. W., Center, T. D. & Driver, F. (2000). Evidence for a new *Cyrtobagous* species (Coleoptera: Curculionidae) on *Salvinia minima* Baker in Florida. *Southwest Entomol* **25**, 299-301.

Hachtel, W. (1996). DNA and gene expression in nonphotosynthetic plastids. In: Handbook of photosynthesis, pp. 349-355. Edited by M. Pessarakli. Marcel Dekker, New York.

Hembree, S. C. (1979). Preliminary reports of some mosquito pathogens from Thailand. *Mosq News* **39**, 575-582.

Hembree, S. C. (1981). Evaluation of the microbial control potential of a *Helicosporidium* sp. (Protozoa: Helicosporida) from *Aedes aegypti* and *Culex quinquefasciatus* from Thailand. *Mosq News* **41**, 770-783.

Higashiyama, T. & Yamada, T. (1991). Electrophoretic karyotyping and chromosomal gene mapping of *Chlorella*. *Nuc Acids Res* **19,** 6191-6195.

Huss, V.A.R., Frank, C., Hartmann, E.C., Hirmer, M., Kloboucek, A., Seidel, B.M., Wenzeler, P. & Kessler, E. (1999). Biochemical taxonomy and molecular phylogeny of the genus *Chlorella* sensu lato (Chlorophyta). *J Phycol* **35,** 587-598.

Kathir, P., LaVoie, M., Brazelton, W.J., Haas, N.A., Lefebvre, P.A & Silflow, C.D. (2003). Molecular map of the *Chlamydomonas reinhardtii* nuclear genome. *Euk Cell* **2,** 362-379.

Keeling, P.J. (2003). Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of Microsporidia. *Fung Genet Biol* **38,** 298-309.

Keeling, P. J. & Fast, N. M. (2002). Microsporidia: Biology and evolution of highly reduced intracellular parasites. *Ann Rev Microbiol* **56**, 93-116.

Keilin, D. (1921). On the life history of *Helicosporidium parasiticum* n.g. sp., a new species of protist parasite in the larvae of *Dashelaea obscura* Winn (Diptera: Ceratopogonidae) and in some other arthropods. *Parasitol* **13,** 97-113.

Kellen, W. R. & Lindegren, J. E. (1973). New host records for *Helicosporidium parasiticum*. *J Invertebr Pathol* **22,** 296-297.

Kellen, W. R. & Lindegren, J. E. (1974). Life cycle of *Helicosporidium parasiticum* in the navelworm *Paramyelois transitella*. *J Invertebr Pathol* **23,** 202-208.

Kim, S.S. & Avery, S.W. (1986). Effects of *Helicosporidium* sp. infection on larval mortality, adult longevity, and fecundity of *Culex salinarius* Coq. *Korean J Entomol* **16,** 153-156.

Knauf, U. & Hachtel, W. (2002). The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Mol Genet Genomics* **267**, 492-497.

Kudo, R. R. (1931). Handbook of protozoology. Thomas, Springfield, Illinois.

Kurtzman, C. P. & Robnett, C. J. (1997). Identification of clinically important ascomycetous yeasts based on the nucleotide divergence in the 5' end of the large-subunit (26S) ribosomal DNA gene. *J Clin Microbiol* **35,** 1216-1223.

Lang, N.J. (1963). Electron-Microscopic demonstration of plastids in *Polytoma*. *J Protozool* **10**, 333-339.

Lee, J.J., Leedale, G.F. & Bradbury, P. (2002). Illustrated guide to the protozoa 2nd Edition. (groups classically considered protozoa and newly discoved ones). Society of Protozoologists, Lawrence, Kansas.

Lindegren, J. E & Hoffman, D. F. (1976). Ultrastructure of some developmental stages of *Helicosporidium* sp. in the navel orangeworm *Paramyelois transitella*. *J Invertebr Pathol* **27,** 105-113.

Lipscomb, D. L., Farris, J. S., Källersjö, M. & Tehler, A. (1998). Support, ribosomal sequences and the phylogeny of the eukaryotes. *Cladistics* **14,** 303-338.

Lockhart, P. J., Steel, M. A. & Penny, D. (1994). Recovering the correct tree under a more realistic model of evolution. *Mol Biol Evol* **11,** 605-612.

Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, Jr, C. T., Saxman, P. R., Stredwick, J. M., Garrity, G. M., Li, B., Olsen, G. J., Pramanik, S., Schmidt, T. M. & Tiedje, J. M. (2000). The RDP (Ribosomal Database Project) continues. *Nucl Acid Res* **28,** 173-174.

Maleszka, R. (1993). Electrophoretic analysis of the nuclear and organellar genomes in the ultra-small alga *Cyanidioschyzon merolae*. *Curr Genet* **24,** 548-550.

Melville, P.A., Benites, N.R., Sinhorini, I.L. & Costa, E.O. (2002). Susceptibility and features of the ultrastructure of *Prototheca zopfii* following exposure to copper sulfate, silver nitrate and chlorexidine. *Mycopathologia* **156,** 1-7.

Mohabeer, A. J., Kaplan, P. J, Southern, Jr, P. M. & Gander, R. M. (1997). Algaemia due to *Prototheca wickerhamii* in a patient with Myasthenia Gravis. *J Clin Microbiol* **35,** 3305-3307.

Moran, N.A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108,** 583-586.

Morell, V. (1996). TreeBASE: the roots of phylogeny. *Science* **273**, 569.

Nedelcu, A. M., Lee, R. W., Lemieux, C., Gray, M. W. & Burger, G. (2000). The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome* **10,** 819-831.

Nedelcu, A. M. (2001). Complex pattern of plastid 16S rRNA gene evolution in nonphotosynthetic green algae. *J Mol Evol* **53**, 670-679

Pekkarinen, M. (1993). Bucephalid trematode sporocysts in brackish-water *Mytilus edulis*, new host of a *Helicosporidium* sp. (Protozoa: Helicosporida). *J Invertebr Pathol* **61**, 214-216.

Perez-Martinez, X., Vasquez-Acevedo, M., Tolkunova, E., Funes, S., Claros, M.G., Davidson, E., King, M.P. & Gonzalez-Halphen, D. (2000). Unusual location of a mitochondrial gene. Subunit III of cytochrome c oxidase is encoded in the nucleus of chlamydomonad algae. *J Biol Chem* **275,** 30144-30152.

Philippe, H. & Adoutte, A. (1998). The molecular phylogeny of Eukaryota: solid facts and uncertainties. In: Evolutionary relationships among Protozoa, pp. 25-57. Edited by G.H. Coombs, K. Vickerman, M.A. Sleigh & A. Warren. Kluwer Academic Publishers.

Pore, R.S. (1985). *Prototheca* taxonomy. *Mycopathologia* **90,** 129-139.

Purrini, K. (1984). Light and electron microscope studies on *Helicosporidium* sp. parasitizing orbitid mites (Oribatei, Acarini) and collembola (Apterygota: Insecta) in forest soils. *J Invertebr Pathol* **44,** 18-27.

Sayre, R. M. & Clark, T. B. (1978). *Daphinia magna* (Cladocera: Chydoroidea) A new host of a *Helicosporidium* sp. (Protozoa: Helicosporida). *J Invertebr Pathol* **31,** 260-261.

Screen, S.E. & St. Leger, R.J. (2000). Cloning, expression, and substrate specificity of a fungal chymotrysin. Evidence for lateral gene transfer from an actinomycete bacterium. *J Bio Chem* **275,** 6689-6694.

Seif, A. I. & Rifaat, M.A. (2001). Laboratory evaluation of a *Helicosporidium* sp. (Protozoa: Helicosporidia) as an agent for the microbial control of mosquitoes. *J Egypt Soc Parasitol* **31**, 21-35.

Sekiguchi, H., Moriya, M., Nakayama, T. & Inouye, I. (2002). Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infosium* (Dictyochophyceae). *Protist* **153,** 157-167.

Shrager, J., Hauser, C., Chang, C.-W., Harris, E.H., Davies, J., McDermott, J., Tamse, R., Zhang, Z. & Grossman, A.R. (2003). *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information. *Plant Physiol* **131,** 401-408.

Simpson, A.G.B. & Roger, A.J. (2002). Eukaryotic evolution: getting to the root of the problem. *Curr Biol* **12,** R691-R693.

Siu, C., Swift, H. & Chiang, K. (1976). Characterization of cytoplasmic and nuclear genomes in the colorless alga *Polytoma*. I. Ultrastructural analysis of organelles. *J Cell Biol* **69**, 352-370.

St. Leger, R.J., Frank, D.C., Roberts, D.W. & Staples, R.C. (1992). Molecular cloning and regulatory analysis of the cuticle-degrading protease structural gene from the entomopathogenic fungus *Metarhizium anisopliae*. *Eur J Biochem* **204,** 991-1001.

Stoebe, B. & Kowallik, K.V. (1999). Gene-cluster analysis in chloroplast genomics. *Trends Genet* **15,** 344-347.

Swofford, D. L. (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tanada, Y. & Kaya, H.K. (1993). Insect pathology. Academic Press, San Diego.

Taylor, F. J. R. (1999). Ultrastructure as a control for protistan molecular phylogeny. *Am Nat* **154 (supplement),** S125-S135.

Tehler, A., Farris, J. S., Lipscomb, D. L. & Källersjö, M. (2000). Phylogenetic analysis of the fungi based on large rDNA data sets. *Mycologia* **92,** 459-474.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acid Res* **24,** 4876-4882.

Turmel, M., Otis, C. & Lemieux, C. (1999). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proc Natl Acad Sci USA* **96**, 10248-10253.

Ueno, R., Urano, N. & Suzuki, M. (2003). Phylogeny of the non-photosynthetic green micro-algal genus *Prototheca* (Trebouxiophyceae, Chlorophyta) and related taxa inferred from SSU and LSU ribosomal DNA partial sequence data. *FEMS Microbiol Lett* **223,** 275-280.

Undeen, A.H. & Vavra, J. (1997). Research methods for entomopathogenic protozoa. In: Manual of techniques in insect pathology, pp. 117-151. edited by L. Lacey. Biological techniques series, Academic Press, San Diego.

Vivares, C.P.,, Gouy, M., Thomarat, F. & Metenier, G. (2002). Functional and evolutionary analysis of a eukaryotic parasitic genome. *Curr Op Microbiol* **5,** 499-505.

Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K. & Sugiura, M. (1997). Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc Natl Acad Sci USA* **94,** 5967-5972.

Waller, R.F., Keeling, P.J., Donald, R.G.K., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A.F., Besra, G.S., Roos, D.S. & McFadden, G.I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **95,** 12352-12357.

Weiser, J. (1964). The taxonomic position of *Helicosporidium parasiticum*, Keilin 1924. *J Protozool (supplement)* **11,** 112.

Weiser, J. (1970). *Helicosporidium parasiticum* Keilin infection in the caterpillar of a hepialid moth in Argentina. *J Protozool* **17,** 440-445.

Williams, B.A.P. & Keeling, P.J. (2003). Cryptic organelles in parasitic protists and fungi. *Adv Parasitol* **54,** 9-67.

Wilson, R. J. M. (2002). Progress with parasite plastids. *J Mol Biol* **319**, 257-274.

Wolff, G., Plante, I., Franz Lang, B., Kuck, U. & Burger, G. (1994). Complete sequence of the mitochondrial DNA of the chorophyte alga *Prototheca wickerhamii.* Gene content and gene organization. *J Mol Biol* **237,** 75-86.

BIOGRAPHICAL SKETCH

Aurélien Tartar was born on January 16, 1976, in Lille, France. After successively giving up early vocations as a professional soccer player, rock star, and writer for the *Lonely Planet* travel guides, Aurélien graduated from high school in 1993, and he eventually settled for a career in biological sciences. He entered the Lycee Faidherbe in Lille and, in 1995, he was accepted into the Institut National Agronomique Paris-Grignon, where he obtained an Agronomy Engineer Degree in 1998. Also in 1998, Aurélien completed his MS degree at the Universite Pierre & Marie Curie (Paris VI-Jussieu), Paris, France.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Drion G. Boucias, Chair
Professor of Entomology and Nematology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James E. Maruniak
Associate Professor of Entomology and Nematology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Byron J. Adams
Assistant Professor of Entomology and Nematology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David G. Clark
Associate Professor of Environmental Horticulture

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William G. Farmerie
Assistant Scientist of Biotechnology

This dissertation was submitted to the Graduate Faculty of the College of Agricultural and Life Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 2004

Dean, College of Agricultural and Life Sciences

Dean, Graduate School